

SEARCH REQUEST FORM

U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

Requestor's Name: Khatol Shahman-shah Serial Number: 09747,521

Date: 11/20/01 Phone: 308-8896 Art Unit: 1645

Search Topic: Room # 80-17 Mail Box # 8E12

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

(1) Please search claims 1-30 attached.
Full search including authors (see bib sheet)
Keywords:

B. anthracis
LF (Lethal Factor Protein)
Immunogenic Composition.
PA (Protective Protein)

see attached claims 1-30 and abstract

(2) Please do a Seq Search on

Seq Id # 1
Polynucleotide
Full search and residue 610-2295 Seq Id # 1

Seq Id # 2

Amino Acid

nuc-1,3
prot-2,4

Full search and

1-735
175-735 Seq Id # 4

residues
1-775
9-202

Seq Id

STAFF USE ONLY

Date completed: 12/3/01

Searcher: Sheppane

Terminal time: _____

Elapsed time: _____

CPU time: _____

Total time: _____

Number of Searches: _____

Number of Databases: _____

Search Site

_____ STIC

_____ CM-1

_____ Pre-S.

Type of Search

_____ N.A. Sequence

_____ A.A. Sequence

_____ Structure

_____ Bibliographic

Vendors

_____ IG Suite

_____ STN

_____ Dialog

_____ APS

_____ Geninfo

_____ SDC

_____ DARC/Questel

_____ Other

RECEIVED
NOV 21 2001
STIC

THIS PAGE BLANK (USPTO)

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:49:24 ; Search time 55.64 Seconds
(without alignments)
-110.818 Million cell updates/sec

Title: US-09-747-521-2_COPY_9_282
Perfect score: 1404
Sequence: 1 KVISMSCLVTAITLSPVFI.....RDVLQLYAPEAFNYMDKFNE 274

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_5/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_5/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_5/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_5/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_5/ptodata/2/1aa/PCFUS.COMB.pep.*
6: /cgn2_5/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARY*

Result No.	Query Score	Match Length	Description
------------	-------------	--------------	-------------

-copy_9_282.ra

Db	1	AGGHDGVMHVKKEKNKDEENKQOEHLKEMHIVKIEYKGEAVKKEAE	60
Qy	86	KLEKVPDVLKMYKAIGKIIYVDGDIKHISLEALSEDKKKIKDIYGDALLHEHYV	145
Db	61	KLEKVPDVLKMYKAIGKIIYVDGDIKHISLEALSEDKKKIKDIYGDALLHEHYV	120
Qy	146	AKEGYEPVLVIOSSDVENTEKALNYYEIGKILSRDILSKINQPKFVLNLIKNA	205
Db	121	AKEGYEPVLVIOSSDVENTEKALNYYEIGKILSRDILSKINQPKFVLNLIKNA	180
Qy	206	SDSDGQDLLFTNQLKEHPTDFSEVFLEQNSNEVQEFKAFAYIIEPQHRDVLQYAPEA	265
Db	181	SDSDGQDLLFTNQLKEHPTDFSEVFLEQNSNEVQEFKAFAYIIEPQHRDVLQYAPEA	240
Qy	266	FNYMDKFNE 274	
Db	241	FNYMDKFNE 249	

RESULT 3

PCFUS.COMB.pep.*

RESULT 1
US-08-021-601-6
; Sequence 6, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Nichols, Peter J.
; APPLICANT: Arora, Naveen
; APPLICANT: Singh, Yogendra
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,601
; FILING DATE: 19930212
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-9880
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-021-601-6

Query Match 91.3%; Score 1282; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.6e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 AGGHDGVMHVKKEKNKDEENKQOEHLKEMHIVKIEYKGEAVKKEAE 85

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:49:25 ; Search time 55.64 Seconds
(Without alignments)
308.996 Million cell updates/sec

Title: US-09-747-521-4
Perfect score: 3913
Sequence: 1 MKRRKVLIPALSTIIVS.....TSTNGIKKILIFSKGYEIG 764

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_AA:*
2: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3667	93.7	735	1	US-08-021-601-4
2	3667	93.7	735	1	US-08-082-849B-4
3	3667	93.7	735	5	PCT-US94-01624-4
4	3622.5	92.6	903	1	US-08-021-601-12
5	3622.5	92.6	903	1	US-08-082-849B-12
6	3622.5	92.6	903	5	PCT-US94-01624-12
7	3490	89.2	719	5	US-08-082-849B-31
8	3490	89.2	719	5	PCT-US94-01624-31
9	789.5	20.2	884	1	US-08-471-044-5
10	789.5	20.2	884	2	US-08-471-044-5
11	789.5	20.2	884	2	US-08-463-483A-5
12	789.5	20.2	884	2	US-08-471-046A-5
13	789.5	20.2	884	2	US-08-470-566B-5
14	789.5	20.2	884	3	US-08-469-334-5
15	789.5	20.2	884	3	US-09-300-529-5
16	789.5	20.2	1346	1	US-08-471-044-23
17	789.5	20.2	1346	2	US-08-471-044-23
18	789.5	20.2	1346	2	US-08-463-483A-23
19	789.5	20.2	1346	2	US-08-471-046A-23
20	789.5	20.2	1346	2	US-08-470-566B-23
21	789.5	20.2	1346	3	US-08-469-334-23
22	789.5	20.2	1346	3	US-09-300-529-23
23	787	20.1	881	4	US-08-960-780-32
24	787	20.1	881	4	US-09-073-898-32
25	777	19.9	1338	1	US-08-471-044-50
26	777	19.9	1338	2	US-08-471-044-50
27	777	19.9	1338	2	US-08-463-483A-50

28	777	19.9	1338	2	US-08-471-046A-50	Sequence 50, Appl
29	777	19.9	1338	2	US-08-470-566B-50	Sequence 50, Appl
30	777	19.9	1338	2	US-08-469-334-50	Sequence 50, Appl
31	777	19.9	1338	3	US-09-300-529-50	Sequence 50, Appl
32	776.5	19.8	784	4	US-09-371-913A-7	Sequence 7, Appl
33	775	19.8	852	1	US-08-471-044-36	Sequence 36, Appl
34	775	19.8	852	2	US-08-471-044-36	Sequence 36, Appl
35	775	19.8	852	2	US-08-463-483A-36	Sequence 36, Appl
36	775	19.8	852	2	US-08-470-566B-36	Sequence 36, Appl
37	775	19.8	852	2	US-08-469-334-36	Sequence 36, Appl
38	775	19.8	852	2	US-09-300-529-36	Sequence 36, Appl
39	775	19.8	852	1	US-08-471-044-21	Sequence 21, Appl
40	766	19.6	834	2	US-08-463-483A-21	Sequence 21, Appl
41	766	19.6	834	2	US-08-471-044-21	Sequence 21, Appl
42	766	19.6	834	2	US-08-463-483A-21	Sequence 21, Appl
43	766	19.6	834	2	US-08-470-566B-21	Sequence 21, Appl
44	766	19.6	834	2	US-08-469-334-21	Sequence 21, Appl
45	766	19.6	834	2	US-08-469-334-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-08-021-601-4
Sequence 4, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Aurora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Needie & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-4

Query Match 93.7%; Score 3667; DB 1; Length 735;
Best local Similarity 97.6%; Pred. No. 2.6e-261;
Matches 717; Conservative 2; Mismatches 16; Indels 0; Gaps 0;
30 EVQENKLNESSSGGLGTFSDINFPQAPMYVTSITGDLSIPSELENIPSENOYF 89

```
|||||
Db 1 EVKQENRLNLSSESSQGLGYYFSDLNFOAPMVYTSSTGDLSPSSELEIPSENOYF 60
QY 90 QSAIWSGFIKYKKSDEYFATISADNHVTMWVDDQEVINKASNSNRIREKGLYQIKIY 149
   |||||||
Db 61 QSAIWSGFIKYKKSDEYFATISADNHVTMWVDDQEVINKASNSNRIREKGLYQIKIY 120
QY 150 QRENTEKGLDFKLWTDSQNKKEVISSDNQLPELKOKSSNRKRSTSGAPYVDDDN 209
   |||||||
Db 121 QRENTEKGLDFKLWTDSQNKKEVISSDNQLPELKOKSSNRKRSTSGAPYVDDDN 180
QY 210 DGIPDSLEVEGYTVYVKKRFTLSPWISNIHEKKGLTYKSSPEKWSASDPYDFEYV 269
   |||||||
Db 181 DGIPDSLEVEGYTVYVKKRFTLSPWISNIHEKKGLTYKSSPEKWSASDPYDFEYV 240
QY 270 GRIDKNSPEARHPLVAAYPIYHVMENIILSKNDOSTQNTDSETRISKNTSRHT 329
   |||||||
Db 241 GRIDKNSPEARHPLVAAYPIYHVMENIILSKNDOSTQNTDSETRISKNTSRHT 300
QY 330 SEVHGNAEVHANTSTSRHTSEVHGNAEYHAYADHSLSLAGEERTWAGLNTADTARL 389
   |||||||
Db 301 SEVHGNAEVHANTSTSRHTSEVHGNAEYHAYADHSLSLAGEERTWAGLNTADTARL 360
QY 390 NANIRYVNTGAPRIYVNLPTTSLVIGKNQTLATIKAKENQSLIAPNNYPSKMLADIA 449
   |||||||
Db 361 NANIRYVNTGAPRIYVNLPTTSLVIGKNQTLATIKAKENQSLIAPNNYPSKMLADIA 420
QY 450 LNAODFSSPTITMNYNOFLELEKTQQLRLDQYGYGNATYFNFGNRYVDGNSMSEV 509
   |||||||
Db 421 LNAODFSSPTITMNYNOFLELEKTQQLRLDQYGYGNATYFNFGNRYVDGNSMSEV 480
QY 510 LPQIETTARIIFNGKDLNVERRIAANVPSDPLETTKPDMLKALKIAGFNPNGNL 569
   |||||||
Db 481 LPQIETTARIIFNGKDLNVERRIAANVPSDPLETTKPDMLKALKIAGFNPNGNL 540
QY 570 QYOGKDITEFDENFDQOQSNIKNQLAELNATNITYVLDKIKLNKMNILIRDKRFHYDR 629
   |||||||
Db 541 QYOGKDITEFDENFDQOQSNIKNQLAELNATNITYVLDKIKLNKMNILIRDKRFHYDR 600
QY 630 NNIAVGADESYYKKAHREYINSTEGLLNTDKDIRKILSGYVEIEDTEGLKEVIYNDY 689
   |||||||
Db 601 NNIAVGADESYYKKAHREYINSTEGLLNTDKDIRKILSGYVEIEDTEGLKEVIYNDY 660
QY 650 DMLNSSLROGDKTIDRKYNDKLPYISNPNKYVNYAYVAKENTINPSNGSTSNIG 749
   |||||||
Db 661 DMLNSSLROGDKTIDRKYNDKLPYISNPNKYVNYAYVAKENTINPSNGSTSNIG 720
QY 750 IKKILIFSKKGEIG 764
   |||||||
Db 721 IKKILIFSKKGEIG 735

RESULT 2
US-08-082-849B-4
: Sequence 4, Application US/08082849B
: Patent No. 5677274
: GENERAL INFORMATION:
: APPLICANT: Leppla, Stephen H.
: APPLICANT: Klimpel, Kurt R.
: APPLICANT: Aroa, Naveen
: APPLICANT: Singh, Yogendra
: APPLICANT: Nichols, Peter J.
: TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
: NUMBER OF SEQUENCES: 35
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
```

```
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-4

Query Match 93.7%; Score 3667; DB 1; Length 735;
Best Local Similarity 97.6%; Pred. No. 2.6e-261;
Matches 717; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

30 EVKQENRLNLSSESSQGLGYYFSDLNFOAPMVYTSSTGDLSPSSELEIPSENOYF 89
1 EVKQENRLNLSSESSQGLGYYFSDLNFOAPMVYTSSTGDLSPSSELEIPSENOYF 60
90 QSAIWSGFIKYKKSDEYFATISADNHVTMWVDDQEVINKASNSNRIREKGLYQIKIY 149
61 QSAIWSGFIKYKKSDEYFATISADNHVTMWVDDQEVINKASNSNRIREKGLYQIKIY 120
150 QRENTEKGLDFKLWTDSQNKKEVISSDNQLPELKOKSSNRKRSTSGAPYVDDDN 209
121 QRENTEKGLDFKLWTDSQNKKEVISSDNQLPELKOKSSNRKRSTSGAPYVDDDN 180
210 DGIPDSLEVEGYTVYVKKRFTLSPWISNIHEKKGLTYKSSPEKWSASDPYDFEYV 269
181 DGIPDSLEVEGYTVYVKKRFTLSPWISNIHEKKGLTYKSSPEKWSASDPYDFEYV 240
270 GRIDKNSPEARHPLVAAYPIYHVMENIILSKNDOSTQNTDSETRISKNTSRHT 329
241 GRIDKNSPEARHPLVAAYPIYHVMENIILSKNDOSTQNTDSETRISKNTSRHT 300
330 SEVHGNAEVHANTSTSRHTSEVHGNAEYHAYADHSLSLAGEERTWAGLNTADTARL 389
301 SEVHGNAEVHANTSTSRHTSEVHGNAEYHAYADHSLSLAGEERTWAGLNTADTARL 360
390 NANIRYVNTGAPRIYVNLPTTSLVIGKNQTLATIKAKENQSLIAPNNYPSKMLADIA 449
361 NANIRYVNTGAPRIYVNLPTTSLVIGKNQTLATIKAKENQSLIAPNNYPSKMLADIA 420
450 LNAODFSSPTITMNYNOFLELEKTQQLRLDQYGYGNATYFNFGNRYVDGNSMSEV 509
421 LNAODFSSPTITMNYNOFLELEKTQQLRLDQYGYGNATYFNFGNRYVDGNSMSEV 480
510 LPQIETTARIIFNGKDLNVERRIAANVPSDPLETTKPDMLKALKIAGFNPNGNL 569
481 LPQIETTARIIFNGKDLNVERRIAANVPSDPLETTKPDMLKALKIAGFNPNGNL 540
570 QYOGKDITEFDENFDQOQSNIKNQLAELNATNITYVLDKIKLNKMNILIRDKRFHYDR 629
541 QYOGKDITEFDENFDQOQSNIKNQLAELNATNITYVLDKIKLNKMNILIRDKRFHYDR 600
630 NNIAVGADESYYKKAHREYINSTEGLLNTDKDIRKILSGYVEIEDTEGLKEVIYNDY 689
```

Db 601 NNIAVGADESVYKHAHREVINSSTEGLLNIDKIRKLISGYIEIDTEGLKEVINDRY 660
QY 690 DMLNSSLRODGKTFIDFKRYNDKPLIYISNPNKVYVAVTKNTIINPSENDTSTNG 749
Db 661 DMLNSSLRODGKTFIDFKRYNDKPLIYISNPNKVYVAVTKNTIINPSENDTSTNG 720
QY 750 IKKILIFSKKGYEIG 764
Db 721 IKKILIFSKKGYEIG 735

RESULT 3
PCT-US94-01624-4
Sequence 4, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Kilmpel, Kurt R.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
STREET: Stewart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-4

Query Match 93.74; Score 3667; DB 5; Length 735;
Best Local Similarity 97.64; Pred. No. 2,6e-261;
Matches 717; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 30 EVKQENRLNSESSESSGGLGYFSDNLFQAPMVVSTSTGDLSPSELENIPSENOYF 89
Db 1 EVKQENRLNSESSESSGGLGYFSDNLFQAPMVVSTSTGDLSPSELENIPSENOYF 60
QY 90 QSAIWSGFIVKKSDEYTFATSDNHNVTMVDQEVINKASNSKIRLEKGRLYQIKIY 149
Db 61 QSAIWSGFIVKKSDEYTFATSDNHNVTMVDQEVINKASNSKIRLEKGRLYQIKIY 120
QY 150 QRENPTKGLDFKLYWDSOKKKEVYSSDNQLPELKOKSSNSKRRKSTSGPYPPDDN 209
Db 121 QRENPTKGLDFKLYWDSOKKKEVYSSDNQLPELKOKSSNSKRRKSTSGPYPPDDN 180
QY 210 DGIPLSLEVEGYIVDVKNKRTFLSPWISNIHEKGLTFYKSSPEKWSYASDPYDFEYVT 269

Db 181 DGIPLSLEVEGYIVDVKNKRTFLSPWISNIHEKGLTFYKSSPEKWSYASDPYDFEYVT 240
QY 270 GRIDKNVSPEARHPLVAAPYIVHVMENIILSKNEDOSTONTDESTRISKNTSTSRHT 329
Db 241 GRIDKNVSPEARHPLVAAPYIVHVMENIILSKNEDOSTONTDESTRISKNTSTSRHT 300
QY 330 SEVHGNAEVHANTSTSRHTSEVHGNAEVHVAIDHSLSLAGERMWTMLNADTARL 389
Db 301 SEVHGNAEVHASFDPDGGSVAGFSNSSTVALDHSLSLAGERMWTMLNADTARL 360
QY 390 NANIRYVNTGAPYINVLPTTSILVYGNQTLATIKAKENQLSQILAPNNYPSKRLAPIA 449
Db 361 NANIRYVNTGAPYINVLPTTSILVYGNQTLATIKAKENQLSQILAPNNYPSKRLAPIA 420
QY 450 LNAODDESSPTITMNYNOFLELEKTKQLRLDTPDOYGYNIATYNEBNGRVAVDTGSMSSEV 509
Db 421 LNAODDESSPTITMNYNOFLELEKTKQLRLDTPDOYGYNIATYNEBNGRVAVDTGSMSSEV 480
QY 510 LPOIETTARIIFNGKDLNVERRIAAYNPSDPLETTKPDMTLEALIKAFGPNPNCNL 569
Db 481 LPOIETTARIIFNGKDLNVERRIAAYNPSDPLETTKPDMTLEALIKAFGPNPNCNL 540
QY 570 QYQKDLITEFDFNFDOQTSQNIKNQLAELNATNTIYVLDKIKLAKNNILIRDRKFHYDR 629
Db 541 QYQKDLITEFDFNFDOQTSQNIKNQLAELNATNTIYVLDKIKLAKNNILIRDRKFHYDR 600
QY 630 NNIAVGADESVYKHAHREVINSSTEGLLNIDKIRKLISGYIEIDTEGLKEVINDRY 689
Db 601 NNIAVGADESVYKHAHREVINSSTEGLLNIDKIRKLISGYIEIDTEGLKEVINDRY 660
QY 690 DMLNSSLRODGKTFIDFKRYNDKPLIYISNPNKVYVAVTKNTIINPSENDTSTNG 749
Db 661 DMLNSSLRODGKTFIDFKRYNDKPLIYISNPNKVYVAVTKNTIINPSENDTSTNG 720
QY 750 IKKILIFSKKGYEIG 764
Db 721 IKKILIFSKKGYEIG 735

RESULT 4
US-08-021-601-12
Sequence 12, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Kilmpel, Kurt R.
APPLICANT: Arora, Naveen
APPLICANT: Nichols, Peter J.
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratl, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414,057

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 404/688-0770
 TELEFAX: 404/688-9880
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 903 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-021-601-12

Query Match 92.6%; Score 3622.5; DB 1; Length 903;
 Best Local Similarity 96.7%; Pred. No. 6.5e-258;

Matches 710; Conservative 3; Mismatches 18; Indels 3; Gaps 1;

```

QY 30 EVKQENRLNSESSESSQGLGYFFSDLNFOAPMVVTSSTTGDLSPSSLENIPISENOYF 89
DB 1 EVKQENRLNSESSESSQGLGYFFSDLNFOAPMVVTSSTTGDLSPSSLENIPISENOYF 60
QY 90 QSAIWSGFIKVKKSDEYTFATSDADNHVTMWDQEVINKASNSNKIRLEKGLYQIKIY 149
DB 61 QSAIWSGFIKVKKSDEYTFATSDADNHVTMWDQEVINKASNSNKIRLEKGLYQIKIY 120
QY 150 QRENPTKGLDFKLYWTDSONKKEVISSDNLOLPELKOKSSNSRRKRSAGPTVPDRDN 209
DB 121 QRENPTKGLDFKLYWTDSONKKEVISSDNLOLPELKOKSSNSRRKRSAGPTVPDRDN 180
QY 210 DGIPOSLEVEGYTVDYKKNRFTLSPWISNIHEKKGLTKYKSSPEKSTASDPYSDPEKYT 269
DB 181 DGIPOSLEVEGYTVDYKKNRFTLSPWISNIHEKKGLTKYKSSPEKSTASDPYSDPEKYT 240
QY 270 GRIDKNVSPPEARHPLVAAPIYHVDMENTILSKNEDOSTQNTDSEFRTISKNTSRTHT 329
DB 241 GRIDKNVSPPEARHPLVAAPIYHVDMENTILSKNEDOSTQNTDSEFRTISKNTSRTHT 300
QY 330 SEVHGNAEYHANTSTSRHTSEVHGNAEYHAVAIDHSLSLAGERTVAETMGLTADTARL 389
DB 301 SEVHGNAEYHASFPIDGVSAGFSNSNSTYVAIDHSLSLAGERTVAETMGLTADTARL 360
QY 390 NANIRYVNTGTAPIYVNLPTTSLVLGKNOTLATIKAKENQLSQILAPNNYPSKNLAPIA 449
DB 361 NANIRYVNTGTAPIYVNLPTTSLVLGKNOTLATIKAKENQLSQILAPNNYPSKNLAPIA 420
QY 450 LNAODDFSSPTITMANYNOFLLEKTKQLRLDQVYGNATYVNFENGRRVVDGSMSEY 509
DB 421 LNAODDFSSPTITMANYNOFLLEKTKQLRLDQVYGNATYVNFENGRRVVDGSMSEY 480
QY 510 LPOIOETTRITIFNGKDLNVERRIAANVPSDPLETTKPDMLKEALKTAFGNEPENGML 569
DB 481 LPOIOETTRITIFNGKDLNVERRIAANVPSDPLETTKPDMLKEALKTAFGNEPENGML 540
QY 570 QYQKGDITEFDENFDQOTSQNTIKNOLAELNATNIYTVLDKIKLMAKMNILIDKRRHYDR 629
DB 541 QYQKGDITEFDENFDQOTSQNTIKNOLAELNATNIYTVLDKIKLMAKMNILIDKRRHYDR 600
QY 630 NNIAGADSVYKKAHREYINSSTEGLLNIDKIRKILISGIYVIEDETEGKEVINDRY 689
DB 601 NNIAGADSVYKKAHREYINSSTEGLLNIDKIRKILISGIYVIEDETEGKEVINDRY 660
QY 690 DMLNTSSLRODGKTFIDFKKYNDKLPYVTSNPYKVNVAVKENTINPSENGDSTNG 749
DB 661 DMLNTSSLRODGKTFIDFKKYNDKLPYVTSNPYKVNVAVKENTINPSENGDSTNG 720
QY 750 IKKIL--IFSCKG 760
DB 721 IKKILKKVYLGKKG 734

```

RESULT 5
 US-08-082-849B-12
 Sequence 12, Application US/08082849B
 Patent No. 5677274

GENERAL INFORMATION:
 APPLICANT: Leppla, Stephen H.
 APPLICANT: Klimpel, Kurt R.
 APPLICANT: Arora, Naveen
 APPLICANT: Singh, Yogendra
 APPLICANT: Nichols, Peter J.
 TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
 TITLE OF INVENTION: Related Methods
 NUMBER OF SEQUENCES: 35
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/082,849B
 FILING DATE: 25-JUN-1993
 CLASSIFICATION: 514

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/021,601

FILING DATE: 12-FEB-1993

ATTORNEY/AGENT INFORMATION:

NAME: Weber, Kenneth A.

REGISTRATION NUMBER: 31,677

REFERENCE/DOCKET NUMBER: 15280-161-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 903 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-082-849B-12

Query Match 92.6%; Score 3622.5; DB 1; Length 903;
 Best Local Similarity 96.7%; Pred. No. 6.5e-258;
 Matches 710; Conservative 3; Mismatches 18; Indels 3; Gaps 1;

```

QY 30 EVKQENRLNSESSESSQGLGYFFSDLNFOAPMVVTSSTTGDLSPSSLENIPISENOYF 89
DB 1 EVKQENRLNSESSESSQGLGYFFSDLNFOAPMVVTSSTTGDLSPSSLENIPISENOYF 60
QY 90 QSAIWSGFIKVKKSDEYTFATSDADNHVTMWDQEVINKASNSNKIRLEKGLYQIKIY 149
DB 61 QSAIWSGFIKVKKSDEYTFATSDADNHVTMWDQEVINKASNSNKIRLEKGLYQIKIY 120
QY 150 QRENPTKGLDFKLYWTDSONKKEVISSDNLOLPELKOKSSNSRRKRSAGPTVPDRDN 209
DB 121 QRENPTKGLDFKLYWTDSONKKEVISSDNLOLPELKOKSSNSRRKRSAGPTVPDRDN 180
QY 210 DGIPOSLEVEGYTVDYKKNRFTLSPWISNIHEKKGLTKYKSSPEKSTASDPYSDPEKYT 269
DB 181 DGIPOSLEVEGYTVDYKKNRFTLSPWISNIHEKKGLTKYKSSPEKSTASDPYSDPEKYT 240
QY 270 GRIDKNVSPPEARHPLVAAPIYHVDMENTILSKNEDOSTQNTDSEFRTISKNTSRTHT 329
DB 241 GRIDKNVSPPEARHPLVAAPIYHVDMENTILSKNEDOSTQNTDSEFRTISKNTSRTHT 300
QY 330 SEVHGNAEYHANTSTSRHTSEVHGNAEYHAVAIDHSLSLAGERTVAETMGLTADTARL 389
DB 301 SEVHGNAEYHASFPIDGVSAGFSNSNSTYVAIDHSLSLAGERTVAETMGLTADTARL 360
QY 390 NANIRYVNTGTAPIYVNLPTTSLVLGKNOTLATIKAKENQLSQILAPNNYPSKNLAPIA 449

```

```

Db 361 NANRYVNTGTADYVNLPTTSLVIGKNGTLATIKAKENQSLIAPNNYPSKNLAPIA 420
QY 450 LNAODDESSPTITMNNYNOFLEKTKQLRLDQOVYGNINATYNNENGRVAVDTSSNSEV 509
Db 421 LNAODDESSPTITMNNYNOFLEKTKQLRLDQOVYGNINATYNNENGRVAVDTSSNSEV 480
QY 510 LPOIETTARIIFNGKDLNVERRIAIVNPSDPLETTKPMPTLKEALKIAGFNEPNCNL 569
Db 481 LPOIETTARIIFNGKDLNVERRIAIVNPSDPLETTKPMPTLKEALKIAGFNEPNCNL 540
QY 570 QYQKDIETEDFNFDOQTSONIKNOELNATNITYVLDRKIKLAKNNILIRDKRFHYDR 629
Db 541 QYQKDIETEDFNFDOQTSONIKNOELNATNITYVLDRKIKLAKNNILIRDKRFHYDR 600
QY 630 NNIAVGADESVYKAEHREVINSSTEGLLNIDNDIRKILSGYIEIEDTEGLKEVINDRY 689
Db 601 NNIAVGADESVYKAEHREVINSSTEGLLNIDNDIRKILSGYIEIEDTEGLKEVINDRY 660
QY 690 DMLNSSLRODGKTFIDFKKYNDKPLIYSNPYKVVNYAVTKENTIIINPSENGDTSTNG 749
Db 661 DMLNSSLRODGKTFIDFKKYNDKPLIYSNPYKVVNYAVTKENTIIINPSENGDTSTNG 720
QY 750 IKKIL--IFSCKG 760
Db 721 IKKILKKVVLGCKG 734

```

RESULT 6
PCT-US94-01624-12
Sequence 12, Application PC/TUS9401624

GENERAL INFORMATION:

APPLICANT: Leppia, Stephen H.
APPLICANT: Kimpel, Kurt R.
APPLICANT: Airoa, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
STREET: Steuart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

PCT-US94-01624-12

```

Query Match 92.6%; Score 3622.5; DB 5; Length 903;
Best Local Similarity 96.7%; Pred. No. 6,5e-258;
Matches 710; Conservative 3; Mismatches 18; Indels 3; Gaps 1;
QY 30 EVKQENRLNSESSESSGGLGYRSDLNFOAPMYVTSSTGDDLSIPSELENIPSENOYF 89
Db 1 EVKQENRLNSESSESSGGLGYRSDLNFOAPMYVTSSTGDDLSIPSELENIPSENOYF 60
QY 90 QSAIWGFIYKKSDDETFATSDNHYTMWDDQEVINKASNSKIRLEKRLQIKIY 149
Db 61 QSAIWGFIYKKSDDETFATSDNHYTMWDDQEVINKASNSKIRLEKRLQIKIY 120
QY 150 QRENPTKGLDFKLYWTDSONKKEVSSDNLQELKOKSSNRKSTSGAPVPPRDN 209
Db 121 QRENPTKGLDFKLYWTDSONKKEVSSDNLQELKOKSSNRKSTSGAPVPPRDN 180
QY 210 DGIPDSLEVEGYVDVKNKRTFLSPWISNHEKGLTKYKSSPKKWSASDPYDFEYVT 269
Db 181 DGIPDSLEVEGYVDVKNKRTFLSPWISNHEKGLTKYKSSPKKWSASDPYDFEYVT 240
QY 270 GRIDKNVSPKARHPLVAAYPYVHYDMENILSKNEDOSTQNTDSEPTISKNTSRTHT 329
Db 241 GRIDKNVSPKARHPLVAAYPYVHYDMENILSKNEDOSTQNTDSEPTISKNTSRTHT 300
QY 330 SEVHGNAEVRHANTSTSRHTSEVHGNAEVRHAVAIDHSLSLGERTMAETMGLANTADTARL 389
Db 301 SEVHGNAEVRHANTSTSRHTSEVHGNAEVRHAVAIDHSLSLGERTMAETMGLANTADTARL 360
QY 390 NANRYVNTGTADYVNLPTTSLVIGKNGTLATIKAKENQSLIAPNNYPSKNLAPIA 449
Db 361 NANRYVNTGTADYVNLPTTSLVIGKNGTLATIKAKENQSLIAPNNYPSKNLAPIA 420
QY 450 LNAODDESSPTITMNNYNOFLEKTKQLRLDQOVYGNINATYNNENGRVAVDTSSNSEV 509
Db 421 LNAODDESSPTITMNNYNOFLEKTKQLRLDQOVYGNINATYNNENGRVAVDTSSNSEV 480
QY 510 LPOIETTARIIFNGKDLNVERRIAIVNPSDPLETTKPMPTLKEALKIAGFNEPNCNL 569
Db 481 LPOIETTARIIFNGKDLNVERRIAIVNPSDPLETTKPMPTLKEALKIAGFNEPNCNL 540
QY 570 QYQKDIETEDFNFDOQTSONIKNOELNATNITYVLDRKIKLAKNNILIRDKRFHYDR 629
Db 541 QYQKDIETEDFNFDOQTSONIKNOELNATNITYVLDRKIKLAKNNILIRDKRFHYDR 600
QY 630 NNIAVGADESVYKAEHREVINSSTEGLLNIDNDIRKILSGYIEIEDTEGLKEVINDRY 689
Db 601 NNIAVGADESVYKAEHREVINSSTEGLLNIDNDIRKILSGYIEIEDTEGLKEVINDRY 660
QY 690 DMLNSSLRODGKTFIDFKKYNDKPLIYSNPYKVVNYAVTKENTIIINPSENGDTSTNG 749
Db 661 DMLNSSLRODGKTFIDFKKYNDKPLIYSNPYKVVNYAVTKENTIIINPSENGDTSTNG 720
QY 750 IKKIL--IFSCKG 760
Db 721 IKKILKKVVLGCKG 734

```

RESULT 7

US-08-082-849B-31

Sequence 31, Application US/08082849B

Patent No. 5677274

GENERAL INFORMATION:

APPLICANT: Leppia, Stephen H.

APPLICANT: Kimpel, Kurt R.

APPLICANT: Airoa, Naveen

APPLICANT: Singh, Yogendra

APPLICANT: Nichols, Peter J.

TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and

TITLE OF INVENTION: Related Methods

NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-31

Query Match 89.2%; Score 3490; DB 1; Length 719;
Best Local Similarity 93.1%; Pred. No. 2.5e-248;
Matches 688; Conservative 6; Mismatches 21; Indels 24; Gaps 2;

OY 30 EVKQENRLNSESSESSOGLLGYFSDLNFOAPMVYTSSTGDLSPSSLENIIPSENOYF 89
DB 1 EVKQENRLNSESSESSOGLLGYFSDLNFOAPMVYTSSTGDLSPSSLENIIPSENOYF 60
OY 90 QSAIWSGFIKVKKSDEYFATSDAHVMTMWVDDDEVINKASNSNKRIRLEKGRLOYIKIOY 149
DB 61 QSAIWSGFIKVKKSDEYFATSDAHVMTMWVDDDEVINKASNSNKRIRLEKGRLOYIKIOY 120
OY 150 QRENPTTEKGLDKFYWTDSQNKKEVISSDNLQPLKOKSSNS---RRKRSTAGPTVP 205
DB 121 QRENPTTEKGLDKFYWTDSQNKKEVISSDNLQPLKOKSSNSNTATIMQRCNFTLQGPVP 180
OY 206 DRDNGCIDPSLEVEGYTVDVKKRFTFLSPRISNIHEKKGITRYKSSPEKWSASDPYSDF 265
DB 181 DRDNGCIDPSLEVEGYTVDVKKRFTFLSPRISNIHEKKGITRYKSSPEKWSASDPYSDF 240
OY 266 EKVYGRIDKNVSPKARHPLVAAYPIVHYDMENIILSKNEDOSTOITDSETRTISKNTSTS 325
DB 241 EKVYGRIDKNVSPKARHPLVAAYPIVHYDMENIILSKNEDOSTOITDSETRTISKNTSTS 300
OY 326 RTHSEVHGNAEVAHANSTSRTHSEVHGNAEVAHAVALIDHSLSLAGEPTMAETMGLNTAD 385
DB 301 RTHSEVHGNAEVAHANSTSRTHSEVHGNAEVAHAVALIDHSLSLAGEPTMAETMGLNTAD 360
OY 386 TARLANANRYVNTGAPTYNVLPPTSLVIGNKQTLATIKAKENOSQILAPNNYPSKNL 445
DB 361 TARLANANRYVNTGAPTYNVLPPTSLVIGNKQTLATIKAKENOSQILAPNNYPSKNL 420
OY 446 APALNAODDSSPPTNNYNOFLLEKTKQLRLDTPQVGNIAIYNGENGVRVDTGSN 505
DB 421 APALNAODDSSPPTNNYNOFLLEKTKQLRLDTPQVGNIAIYNGENGVRVDTGSN 460
OY 506 WSEVLPQIOETTARTIINGKDLNVERIAAVNPSDPLETTKPDMTLKEALKIAGFNEP 565
DB 461 WSEVLPQIOETTARTIINGKDLNVERIAAVNPSDPLETTKPDMTLKEALKIAGFNEP 520
OY 566 NGNIQYQCKDITREFDNFNDQOTSQNIKNQDLAELNATNTIYVLDKIKLAKNMILIRDKRF 625

DB 521 NGNIQYQCKDITREFDNFNDQOTSQNIKNQDLAELNATNTIYVLDKIKLAKNMILIRDKRF 580
OY 626 HYDRNNIAGADESVYKAEHREVINSSTEGLLNTDKDIRKILSGYIEIDTEGLKEVI 685
DB 581 HYDRNNIAGADESVYKAEHREVINSSTEGLLNTDKDIRKILSGYIEIDTEGLKEVI 640
OY 666 NDRYMLNLISSLRQDGKTFIDFKRYNDKLPYISPNPKVNVYAVTKENTIIINPSENGDT 745
DB 641 NDRYMLNLISSLRQDGKTFIDFKRYNDKLPYISPNPKVNVYAVTKENTIIINPSENGDT 700
OY 746 STNGIKRILIFSKKGYEIG 764
DB 701 STNGIKRILIFSKKGYEIG 719

RESULT 8
PCT-US94-01624-31
Sequence 31, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Lepola, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arota, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
RELATED METHODS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Stewart Street Tower, 20th Floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-31

Query Match 89.2%; Score 3490; DB 5; Length 719;
Best Local Similarity 93.1%; Pred. No. 2.5e-248;
Matches 688; Conservative 6; Mismatches 21; Indels 24; Gaps 2;

OY 30 EVKQENRLNSESSESSOGLLGYFSDLNFOAPMVYTSSTGDLSPSSLENIIPSENOYF 89
DB 1 EVKQENRLNSESSESSOGLLGYFSDLNFOAPMVYTSSTGDLSPSSLENIIPSENOYF 60
OY 90 QSAIWSGFIKVKKSDEYFATSDAHVMTMWVDDDEVINKASNSNKRIRLEKGRLOYIKIOY 149
DB 61 QSAIWSGFIKVKKSDEYFATSDAHVMTMWVDDDEVINKASNSNKRIRLEKGRLOYIKIOY 120


```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,046A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-471-046A-5

Query Match      20.2%; Score 789.5; DB 2; Length 884;
Best Local Similarity 29.4%; Pred. No. 8.3e-50;
Matches 253; Conservative 134; Mismatches 309; Indels 165; Gaps 34;

QY 1 MKKRVLLPMLALSTLVSS--TGNLEVIQAEVK-----OENRLNSESSESSGGLGY 51
DB 1 MNMKKKLSVVTCTLLAPLMLNGNANAYADSKTQISTTQNO---OKEHDRKGLGY 57
QY 52 YPSDLNFOAPMYVTSSTGDLSPSSSELEN--IPSENYFQSAIWSGFTKVKSDPYTPA 109
DB 56 YFKGKGF--SNLTMFAPTRSTLIYDQGTANKLLDKKQGEYQSIKWTGLQSGTGDFTFN 116
QY 110 TSADNHYTMVDDQEVYINKASNSNKRIRLEKGRLYQIKIORYENPTEKGLD-----FKL 163
DB 117 LSEDEQALIEINGKLIISNKKERQVVAHLEKGLVPIKIEYQSD--TKFNIIDSKTEKELT 174
QY 164 YMTDSQNKKEVISDNQLPBLKOKS-----NSRKKRSTSGPFTVPROND 210
DB 175 FKIDSNQOQOQDELRLRPFENKKESEFLAKPSKINLFTQKMKREIDED---IDTDD 231
QY 211 GIPDSLEVEGYVDVKNKFTPLSPWISNIHERKGLTKYKSSPEKSTASDPYSDEKVTG 270
DB 232 SIPDLMEENGTYI---QRIAVKMDSD--ASKGYTKFVSNPLESHHTVDPTDYDEKAR 286
QY 271 RIDKNSPFAHRHLYAAYIVHVDENILLSKNEDOSITNTSETRTIKNTSTKTHHS 330
DB 287 DLDLSNAKETFPPLVAAPFSVNVSMKYLSPNENLS----- 323
QY 331 EVHGNAEVHANTSTSRTHSEVHGNAEVH-----AVAIHSLSLAGEPTMAETMG-- 380
DB 324 ---NSVESHSSTNMWSTTNEGASVEAGICPKGISFGVSNVYOHSETVADE--WGSTGTNT 378
QY 381 --LNTADTARLANIRYVVTGTAPYIYNVLTPTSLVLGKNOTLATIKAKENQLSQILAPNN 438
DB 379 SQPNTASAGYLVANAYRYNNVNGAIVDAKPTTISFVL--NNDTIATITAKNSATALNISP 437

```

```

QY 439 YPSKNLAPIALMAODDESSPTITMNVNOFLEKTKQILRDTDOVGNINATYFNGRV 498
DB 438 SYPKKGONCIAITSMDFNSHPTILNKKQVDNLLNKKPMMLEETNOTDG---YKIKDTHG 494
QY 499 RYDTSNNSSEVLPQOETTARITFNGKDLNLYERRIAVNPSPDLETTRKPDMTKEALKI 558
DB 495 NIVTGGEMNGVIOQIKAKTASITIVDGE--RAVEKRVAAKDEPNEDKT--PSLTLLDALK 552
QY 559 AF--GFNEPENGMOYQKGDITEFD--NEPDQTSQNIKNQLAEL-----NATNITYLDK 609
DB 553 SYPDEKEKEGLEGLYYKKNKRYESSWMTYIDENYAKVYKQMDITGKRDVSHLDV--- 609
QY 610 IKLAKMNLIPDKRPHYDRNNIAVGADESVYKAEHREYINSSTEG-----LL 657
DB 610 -KLTPKMNVTIK--LSILYDN--AESDMSIGKMTNTNIVSGNNGKQYSSNPDANLT 664
QY 658 LNIID-----KDIRKILSGYIYIEDTE-----GLKEVINDRMDLN----- 693
DB 665 LNTDQOEKLNKRNDYIISLYKSEKNTQCEITIDDEIYIPITTKYVNVKNDYKRLDIIAH 724
QY 694 -----ISSLRQDGKFTIDFKYNDKLPYISNPKYKVVAVYKENTLINPSENGDT--S 746
DB 725 NIKSNPISLSH-----IKTNDIEITLFMDISI--TDVASIKPEN--LTPSEIKQIYS 772
QY 747 TNGIK---KILFSSKKGYEIG 764
DB 773 RYGIKLEDDGILDKKGHNG 793

RESULT 13
US-08-470-566B-5
Sequence 5, Application US/08470566B
Patent No. 5872212
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estrich, Juan J
TITLE OF INVENTION: No. 5872212a1s Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5872212a1s Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,566B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993

```

```
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-566B-5

Query Match      20.2%; Score 789.5; DB 2; Length 884;
Best Local Similarity 29.4%; Pred. No. 8.3e-50;
Matches 253; Conservative 134; Mismatches 309; Indels 165; Gaps 34;

QY 1 MKRRKVLPLMALSTIVSS--TGNLEVIOAEVK-----QENRLNSESSESSQGLGY 51
DB 1 MKMKKKLASVVTCTLLAPMELNGNVAVADSKTNOISTOKNO---QKEMRKGLGY 57
QY 52 YFSDLNFOAMVVTSSTTGDLSTIPSSLEN--IPSENOYFOSAIWSGFIVKKSDEYTPA 109
DB 58 YFKGKDF-SULTMPAPRSDTLIYDQOTANKKLDKKOEYOSIRWIGLISKETGDTFN 116
QY 110 TSADNHTVMVDDOEVIKASNSKIRLEKGRLOYKIOYORENPTKGLD-----FKL 163
DB 117 LSEDEQAIIIEKGIISNKKEKOVHLEKGLVPIKIEYQSD--TFKNIDSKTFKELKL 174
QY 164 YMTDSOKKKEVVISDNQLBELKOKS-----NSRKRKTSAGPYPPDDND 210
DB 175 FKIDSONQPOVOODELRNPEFNKESQOEPLAPSKINLTQKKRIDE--TDTGDD 231
QY 211 GIPDSLEVEGYTVDKKRFPLSPWISNHEKGLITRKYSPEKWSAPSPDFEYTG 270
DB 232 SIFDLMEENYTT---QNIIVAKWDSL-ASKGYTKFVSNPLESHVGDPTDYEKAR 286
QY 271 RIDKNVSPARHPLVAAYPIVHVMENIILSKNEDOSTQNTDSETRTISKNTSRHTS 330
DB 287 DLDLSNAKETFNPLVAAPSVNEMERILSPNENLS----- 323
QY 331 EVMGNAEVNANTSTSRHTSEVHGNAEVH-----AVAIHSLSLAGETMAETWG-- 380
DB 324 ---NSVSHSTWMSYNTGASVEAGIGPKGISFGVSVNYQHSETVAQE--WGTSTGNT 378
QY 381 --LNTADTARLNNIRVNTGTAPIVNVLPPTSLVLGKNQTLATIKKENQLSQILPNN 438
DB 379 SQFTASAGYLNANVRNNVGTGAIYVKPTTSFVL--NNDIATITAKSNSTALNISPE 437
QY 439 YPSKNIAPIALNADODFSPTITMANYNOFLEKTKQLRLDQOVYGNATYFENGVRV 498
DB 438 SYPKKGNGIAITSMDFNSHPITLNNKYVDNLLNNKPMLETNQJG---VYIKDTHG 494
QY 499 RVDTGSWSEVLPJOIQTETARIIFNGKDLNVERRIAANPSDELTTKPMTEKALKI 558
DB 495 NIYVGGMGNGVIOIKAKTASIIYDDGE-RVAEKRAKADVENEDXT-PSLTLLKDALKT 552
QY 559 AF--GFNEPNGNLQYOGKDTEPDF--NPDQOTSQNIKNQALD-----NATNIYTVLDK 609
DB 553 SYPEIKETIEIGLTYKKKPIYESVMTYLDENTAKVEYTKQINDTGGFKVDVSHLYDV--- 609
QY 610 IKLNKAKNIILRDKRFHYDRNNINAVGADESVYKAEHREVINSSTEG-----LL 657
DB 610 -KLTPKKNWVTK-LSLIYDN---AESNDNSIGKWTNTNIVYSGNNGKQOYSSNPNDAFLT 664
QY 658 LNIID-----KDIKIIISGIYVEIDTE-----GLKEYVNDYIDMLN----- 693
DB 665 LNTDAQERLNNRDYIYSLYKSEKNTQCEITIDGETIYPIYTTKTVNWKDKYKRDIIDAH 724
QY 694 -----ISSLRQDGKTFIDFKKYNDKLPVLYISNPNYKVNVAVTKENTIIPSENGDT-S 746
```

```
DB 725 NIKSNPISLSH-----IKTNDIEFLWDDISI-TDVASIKPEN--LTDSKIKOYIS 772
QY 747 TNGIK---KILISKKGYEIG 764
DB 773 RYGIKLEDGILIDKKGGINHYG 793

RESULT 14
US-08-469-334-5
; Sequence 5, Application US/08469334
; Patent No. 5990383
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalin M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5990383el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,334
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/463,483
; FILING DATE:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE:
; APPLICATION NUMBER: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8615
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-469-334-5

Query Match      20.2%; Score 789.5; DB 2; Length 884;
Best Local Similarity 29.4%; Pred. No. 8.3e-50;
Matches 253; Conservative 134; Mismatches 309; Indels 165; Gaps 34;

QY 1 MKRRKVLPLMALSTIVSS--TGNLEVIOAEVK-----QENRLNSESSESSQGLGY 51
||| : : : : :| : : :| : :| :| :| :|
```

```

Db      1 MKNMKKLASVVTCTLLAPMFLNGVNAVYADSKTNOISTOKNO---QKENDRGGLGY 57
Qy      52 YFSDLNFOAPWVYSSTTGDLSPSSSELEN--IPSENOYFOSAIWSGFIKYKKSDEYFA 109
      58 YFKGKDF-SNLTMPAPTRDSTLIYDQOTANKLLDKKQOEYOSIRWIGLQSKETGDTFFN 116
Qy      110 TSADNHVTMWVDDQEVINKASNSNKIRLEKGRLYOIKIOYORENTEKGLD-----FKL 163
      117 LSEDOAIIIEINGKLIISNKGKQOVHLEKGLVPIKIEYOSD--TKFNIDSKTFKEKL 174
Qy      164 YWTDSONKKEVISSDNLQPLPELKQKSS-----NSRKKRSTAGPTVPDRND 210
      175 FKIDSONQPOOVQODELNRPEFNKESQEFLLAKPSKINLFTQKMKREIDED---TDTDGD 231
Qy      211 GIPDSLEVEGYTVDKNKRTEFSPWISNIHEKGLTKYKSSPEKWSSTADSPYDFEKVTG 270
      232 SIPDLMEENGTYI-----QNRIVAKWDDSL-ASKGYTKFVSNPLSHYTGDPYTDYEKAAR 286
Qy      271 RIDKRVSPDEARPIVAAPYIVHDMENIILSKNEDOSTQNTDSETRTTSKNTSRTHTS 330
      287 DIDLNAKETEPPLVAAPPSVAVSMKEVILSPENILS----- 323
Qy      331 EVHGNAEVHANTSTSRTHTEVHGNAEVH-----AVAIHDSLAGERTWAETMG-- 380
      324 ---NSVESHSSTNMSTYTMTEGASVAGIGPKGISFGVSNVYOSHETVAOE--WGTSTGNT 378
Qy      361 --LNTADTARLANIRVYNTGTAPIYVNLPTTSLVGNKQNLATIKARENLSQIILAPNN 438
      379 SOFNFTASAGYLANAVRNNGVGTALYDVKPTTSFVL--NNDTATITAKSNSALNISGE 437
Qy      439 YPSPKNLAPILNADDDSPITMNNYNOFLELEKTQOLRDTDOYGNIAIYNNENGRV 498
      438 STPKGONGIATTSMDRNSHPTLNKKQOVNLLNKKPMLETNOTDG--YVKIKDTHG 494
Qy      499 RVDGSSNMSEVLPOIETTARLIIFNGKDLNLYERIIAANVPSDPLETTKPDWTLKEALKI 558
      495 NIVTGEMNGVYIOIKAKRTASITVDDGE-RVAEKRVAAKDYENPDKT-PSLTLDALKL 552
Qy      559 AF--GFNPNGNLOYOGDITFEFD--NFDQOTSONIKQIAEL-----NATNITYVLDK 609
      553 SYPDRIKIEIGLLYYKKNRPYESSVMTYLDENTAKEVTKQIMDTGKFKDYSHLYDV--- 609
Qy      610 IKLNKKNMILIFDRKRFHNRNIAVGADESUYKEAHRVINSSTEG-----LL 657
      610 -KLTPKAMVYTK-LSILYDN---AESNDNSIGKWTNTNIVSGNGKKOYSSNPDANLT 664
Qy      658 LNID-----KDIRKILSGYIIEIDTE-----GLKEVINDRYMLN----- 693
      665 LNTDQOEKLNKRDYIISLYMKSEKNTQCEITIDGEIYPIITTKYVNVAKNDKRLDIIAH 724
Qy      694 -----ISSLRDQKGTFFIDFKKYNDKLPYISNPNYKAVNAVYAVTKENTIIINSENGDT-S 746
      725 NIKSNPISLSLH-----IKTNDIEITLFWMDISIT-JDVASIKPEN--LTDSEIKQIAYS 772
Qy      747 TNGIK---KILIFSCKGYEIG 764
      773 RYGIKLEDGILIDKKGGIHYG 793

```

```

TITLE OF INVENTION: Genes Encoding Insecticidal Proteins
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESS: No. 606783artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/300,529
FILING DATE: TBA
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/469,334
FILING DATE: 06-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-19506L
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
us-09-300-529-5

Query Match 20.2%, Score 789.5; DB 3; Length 884;
Best Local Similarity 29.4%, Pred. NO. 8.3e-50;
Matches 253; Conservative 134; Mismatches 309; Indels 165; Gaps 34;

Qy      1 MKNMVKLPLMALSTIIIVSS--TGNLEVIOAEVK-----QENRLNSESSSSGILGY 51
      1 MKNMKKLASVVTCTLLAPMFLNGVNAVYADSKTNOISTOKNO---QKENDRGGLGY 57
Db      1 MKNMKKLASVVTCTLLAPMFLNGVNAVYADSKTNOISTOKNO---QKENDRGGLGY 57
Qy      52 YFSDLNFOAPWVYSSTTGDLSPSSSELEN--IPSENOYFOSAIWSGFIKYKKSDEYFA 109
      58 YFKGKDF-SNLTMPAPTRDSTLIYDQOTANKLLDKKQOEYOSIRWIGLQSKETGDTFFN 116
Qy      110 TSADNHVTMWVDDQEVINKASNSNKIRLEKGRLYOIKIOYORENTEKGLD-----FKL 163
      117 LSEDOAIIIEINGKLIISNKGKQOVHLEKGLVPIKIEYOSD--TKFNIDSKTFKEKL 174
Db      117 LSEDOAIIIEINGKLIISNKGKQOVHLEKGLVPIKIEYOSD--TKFNIDSKTFKEKL 174
Qy      164 YWTDSONKKEVISSDNLQPLPELKQKSS-----NSRKKRSTAGPTVPDRND 210
      175 FKIDSONQPOOVQODELNRPEFNKESQEFLLAKPSKINLFTQKMKREIDED---TDTDGD 231
Db      175 FKIDSONQPOOVQODELNRPEFNKESQEFLLAKPSKINLFTQKMKREIDED---TDTDGD 231
Qy      211 GIPDSLEVEGYTVDKNKRTEFSPWISNIHEKGLTKYKSSPEKWSSTADSPYDFEKVTG 270
      232 SIPDLMEENGTYI-----QNRIVAKWDDSL-ASKGYTKFVSNPLSHYTGDPYTDYEKAAR 286

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:49:24 ; Search time 55.64 Seconds
(without alignments)
110.818 Million cell updates/sec

Title: US-09-747-521-2_COPY_9_282
Perfect score: 1404
Sequence: 1 KVISMSCLVTAITLSEPFV1.....RDVLQLYAPEAFNYMDKFNK 274

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues
Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1282	91.3	456	1	US-08-021-601-6
2	1282	91.3	456	1	US-08-082-849B-6
3	1282	91.3	456	5	PCT-US94-01624-6
4	1282	91.3	472	1	US-08-021-601-8
5	1282	91.3	472	1	US-08-082-849B-8
6	1282	91.3	472	5	PCT-US94-01624-8
7	1282	91.3	508	1	US-08-021-601-10
8	1282	91.3	508	1	US-08-082-849B-10
9	1282	91.3	508	5	PCT-US94-01624-10
10	1282	91.3	776	1	US-08-021-601-2
11	1282	91.3	776	1	US-08-082-849B-2
12	1282	91.3	776	6	PCT-US94-01624-2
13	442	91.5	800	6	PCT-US94-01624-2
14	113	8.0	593	1	US-07-961-522-4
15	113	8.0	593	1	US-08-217-438-4
16	113	8.0	593	1	US-08-217-438-5
17	113	8.0	593	1	US-08-467-850A-100
18	113	8.0	593	1	US-08-321-978-4
19	113	8.0	593	2	US-08-710-584-4
20	113	8.0	593	2	US-08-478-435-100
21	113	8.0	593	2	US-08-337-483-100
22	113	8.0	593	2	US-08-478-373-100
23	113	8.0	593	3	US-08-474-671-100
24	113	8.0	593	3	US-08-483-577A-100
25	113	8.0	593	4	US-08-897-438-100
26	112	8.0	1786	4	US-08-973-462-8
27	111	7.9	700	1	US-07-720-589-2

28	111	7.9	700	2	US-08-785-190-2	Sequence 2, Appl
29	111	7.9	700	4	US-08-235-836C-66	Sequence 66, Appl
30	111	7.9	700	5	PCT-US92-05539-2	Sequence 2, Appl
31	111	7.9	1104	4	US-08-923-992A-4	Sequence 4, Appl
32	111	7.9	1164	4	US-08-923-992A-2	Sequence 2, Appl
33	108.5	7.7	937	1	US-08-253-155A-31	Sequence 31, Appl
34	108.5	7.7	1164	4	US-08-923-992A-10	Sequence 10, Appl
35	107	7.6	3111	2	US-08-460-309-4	Sequence 4, Appl
36	107	7.6	3111	2	US-08-125-077-4	Sequence 4, Appl
37	106.5	7.6	381	2	US-08-858-052-3	Sequence 3, Appl
38	106.5	7.6	381	3	US-09-200-284-3	Sequence 3, Appl
39	106.5	7.6	1588	5	PCT-US93-07261-11	Sequence 11, Appl
40	106.5	7.6	1663	5	PCT-US93-07261-16	Sequence 16, Appl
41	106	7.5	976	4	US-09-104-324B-4	Sequence 4, Appl
42	106	7.5	1098	4	US-08-923-992A-8	Sequence 8, Appl
43	105.5	7.5	732	2	US-08-533-669A-18	Sequence 18, Appl
44	102.5	7.3	1128	4	US-08-923-992A-6	Sequence 6, Appl
45	102	7.3	337	1	US-08-445-135-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-021-601-6
Sequence 6, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-9880
TELEFAX: 404/688-0770
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-6

Query Match 91.3%; Score 1282; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.6e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGHGIVGVNHVKEKKNKRNKRKDEARNKTOEHLKEIMKHIVKLEVGAEVKKAEAE 85

Db 1 AGGHGDMHVKKEKKDKDKRDEERNKTOEHLKEIMKHIVKIEVKGEEAVKKEAAE 60
QY 86 KLEKVPDVLKEMKAKGKIYIVDGDITKHISLEALSEDKKKIKDYGKDALHHEHYV 145
Db 61 KLEKVPDVLKEMKAKGKIYIVDGDITKHISLEALSEDKKKIKDYGKDALHHEHYV 120
QY 146 AKEGEPLVLIQSESDYVENTERKALNVEYEGIKLSRDIILSKINOPYOKFLDVLNTIKNA 205
Db 121 AKEGEPLVLIQSESDYVENTERKALNVEYEGIKLSRDIILSKINOPYOKFLDVLNTIKNA 180
QY 206 SDSGODLLFTNOLKEHPTDPSVEFLQNSNEVOEFAKAFAYIIEPOHHDVLIQVAPEA 265
Db 181 SDSGODLLFTNOLKEHPTDPSVEFLQNSNEVOEFAKAFAYIIEPOHHDVLIQVAPEA 240
QY 266 FNYMDKENE 274
Db 241 FNYMDKENE 249

RESULT 2

US-08-082-849B-6
Sequence 6, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Ariora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
TITLE OF INVENTION: Related Methods
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-6

Query Match 91.3%; Score 1282; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 6, 6e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 AGGHGDMHVKKEKKDKDKRDEERNKTOEHLKEIMKHIVKIEVKGEEAVKKEAAE 85

Db 1 AGGHGDMHVKKEKKDKDKRDEERNKTOEHLKEIMKHIVKIEVKGEEAVKKEAAE 60
QY 86 KLEKVPDVLKEMKAKGKIYIVDGDITKHISLEALSEDKKKIKDYGKDALHHEHYV 145
Db 61 KLEKVPDVLKEMKAKGKIYIVDGDITKHISLEALSEDKKKIKDYGKDALHHEHYV 120
QY 146 AKEGEPLVLIQSESDYVENTERKALNVEYEGIKLSRDIILSKINOPYOKFLDVLNTIKNA 205
Db 121 AKEGEPLVLIQSESDYVENTERKALNVEYEGIKLSRDIILSKINOPYOKFLDVLNTIKNA 180
QY 206 SDSGODLLFTNOLKEHPTDPSVEFLQNSNEVOEFAKAFAYIIEPOHHDVLIQVAPEA 265
Db 181 SDSGODLLFTNOLKEHPTDPSVEFLQNSNEVOEFAKAFAYIIEPOHHDVLIQVAPEA 240
QY 266 FNYMDKENE 274
Db 241 FNYMDKENE 249

RESULT 3

PCT-US94-01624-6
Sequence 6, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Ariora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Stewart Street Tower, 20th Floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-6

Query Match 91.3%; Score 1282; DB 5; Length 456;
Best Local Similarity 100.0%; Pred. No. 6, 6e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 AGGHGDMHVKKEKKDKDKRDEERNKTOEHLKEIMKHIVKIEVKGEEAVKKEAAE 85
Db 1 AGGHGDMHVKKEKKDKDKRDEERNKTOEHLKEIMKHIVKIEVKGEEAVKKEAAE 60

QY 86 KLEKVPDVLKMYKAIGKIIYVDGDTTKHISLEALSDKKKIKDIYKDALLHEHYV 145
| | | | |
DB 61 KLEKVPDVLKMYKAIGKIIYVDGDTTKHISLEALSDKKKIKDIYKDALLHEHYV 120
| | | | |
QY 146 AKEGYEPVLVIQSSSEDIYVENTEKALNVYYEIGKILSRDLISKINOPYQKFLDVLNTIKNA 205
| | | | |
DB 121 AKEGYEPVLVIQSSSEDIYVENTEKALNVYYEIGKILSRDLISKINOPYQKFLDVLNTIKNA 180
| | | | |
QY 206 SSDSDODLFTNQLKHPDVSVEFLQNSNEVEYFAKAFAYIIEPQHRDVLQLYAPPA 265
| | | | |
DB 181 SSDSDODLFTNQLKHPDVSVEFLQNSNEVEYFAKAFAYIIEPQHRDVLQLYAPPA 240
| | | | |
QY 266 FNYMDKFNE 274
| | | | |
DB 241 FNYMDKFNE 249
| | | | |

RESULT 4

US-08-021-601-8
; Sequence 8, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Nichols, Peter J.
; APPLICANT: Atora, Naveen
; APPLICANT: Singh, Yogendra
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021.601
; FILING DATE: 19930212
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-021-601-8

Query Match 91.3%; Score 1282; DB 1; Length 472;
Best Local Similarity 100.0%; Pred. No. 6.9e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 AGHGQVGNHNVKEKKNKDEERNKKTQDEHLKEIKKHIVKIEVGEAEVKKKEAAE 85
| | | | |
DB 4 AGHGQVGNHNVKEKKNKDEERNKKTQDEHLKEIKKHIVKIEVGEAEVKKKEAAE 63
| | | | |
QY 86 KLEKVPDVLKMYKAIGKIIYVDGDTTKHISLEALSDKKKIKDIYKDALLHEHYV 145
| | | | |
DB 64 KLEKVPDVLKMYKAIGKIIYVDGDTTKHISLEALSDKKKIKDIYKDALLHEHYV 123
| | | | |

QY 146 AKEGYEPVLVIQSSSEDIYVENTEKALNVYYEIGKILSRDLISKINOPYQKFLDVLNTIKNA 205
| | | | |
DB 124 AKEGYEPVLVIQSSSEDIYVENTEKALNVYYEIGKILSRDLISKINOPYQKFLDVLNTIKNA 183
| | | | |
QY 206 SSDSDODLFTNQLKHPDVSVEFLQNSNEVEYFAKAFAYIIEPQHRDVLQLYAPPA 265
| | | | |
DB 184 SSDSDODLFTNQLKHPDVSVEFLQNSNEVEYFAKAFAYIIEPQHRDVLQLYAPPA 243
| | | | |
QY 266 FNYMDKFNE 274
| | | | |
DB 244 FNYMDKFNE 252
| | | | |

RESULT 5

US-08-082-849B-8
; Sequence 8, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Atora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/082.849B
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/021,601
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-161-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-082-849B-8

Query Match 91.3%; Score 1282; DB 1; Length 472;
Best Local Similarity 100.0%; Pred. No. 6.9e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 AGHGQVGNHNVKEKKNKDEERNKKTQDEHLKEIKKHIVKIEVGEAEVKKKEAAE 85
| | | | |
DB 4 AGHGQVGNHNVKEKKNKDEERNKKTQDEHLKEIKKHIVKIEVGEAEVKKKEAAE 63
| | | | |
QY 86 KLEKVPDVLKMYKAIGKIIYVDGDTTKHISLEALSDKKKIKDIYKDALLHEHYV 145
| | | | |
DB 64 KLEKVPDVLKMYKAIGKIIYVDGDTTKHISLEALSDKKKIKDIYKDALLHEHYV 123
| | | | |

QY 146 AKESYEPVLVIOSESDYVENTEKALNYYEIGKILSRDIISKINOPYOKFLDVLNTITKNA 205
DB 124 AKESYEPVLVIOSESDYVENTEKALNYYEIGKILSRDIISKINOPYOKFLDVLNTITKNA 183
QY 206 SDSGQDLFTNOLKEHPDTSVEFLSEONSNEVOEFAKAFAYTIEPOHRVDLQLYAPEA 265
DB 184 SDSGQDLFTNOLKEHPDTSVEFLSEONSNEVOEFAKAFAYTIEPOHRVDLQLYAPEA 243
QY 266 FNYMDKFNE 274
DB 244 FNYMDKFNE 252

RESULT 6
PCT-US94-01624-8
Sequence 8, Application PC/US9401624

GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
RELATED METHODS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE and CREW
STREET: Stewart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-8

Query Match 91.3%; Score 1282; DB 5; Length 472;
Best Local Similarity 100.0%; Pred. No. 6.9e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGHGVDGMHYKKEKKKKDEKRRKDEERNKTOEHLKEIMKHIVKIEVKGGEAVKKEAAE 85
DB 4 AGHGVDGMHYKKEKKKKDEKRRKDEERNKTOEHLKEIMKHIVKIEVKGGEAVKKEAAE 63
QY 86 KLEKVPDVLVEMTKAIGKTIYVDSGDTKHSLEALSEDKKTKIDYGDALLHEHYV 145
DB 64 KLEKVPDVLVEMTKAIGKTIYVDSGDTKHSLEALSEDKKTKIDYGDALLHEHYV 123
QY 146 AKESYEPVLVIOSESDYVENTEKALNYYEIGKILSRDIISKINOPYOKFLDVLNTITKNA 205

DB 124 AKESYEPVLVIOSESDYVENTEKALNYYEIGKILSRDIISKINOPYOKFLDVLNTITKNA 183
QY 206 SDSGQDLFTNOLKEHPDTSVEFLSEONSNEVOEFAKAFAYTIEPOHRVDLQLYAPEA 265
DB 184 SDSGQDLFTNOLKEHPDTSVEFLSEONSNEVOEFAKAFAYTIEPOHRVDLQLYAPEA 243
QY 266 FNYMDKFNE 274
DB 244 FNYMDKFNE 252

RESULT 7
US-08-021-601-10
Sequence 10, Application US/08021601
Patent No. 5591631

GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 508 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-10

Query Match 91.3%; Score 1282; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 7.5e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGHGVDGMHYKKEKKKKDEKRRKDEERNKTOEHLKEIMKHIVKIEVKGGEAVKKEAAE 85
DB 1 AGHGVDGMHYKKEKKKKDEKRRKDEERNKTOEHLKEIMKHIVKIEVKGGEAVKKEAAE 60
QY 86 KLEKVPDVLVEMTKAIGKTIYVDSGDTKHSLEALSEDKKTKIDYGDALLHEHYV 145
DB 61 KLEKVPDVLVEMTKAIGKTIYVDSGDTKHSLEALSEDKKTKIDYGDALLHEHYV 120
QY 146 AKESYEPVLVIOSESDYVENTEKALNYYEIGKILSRDIISKINOPYOKFLDVLNTITKNA 205
DB 121 AKESYEPVLVIOSESDYVENTEKALNYYEIGKILSRDIISKINOPYOKFLDVLNTITKNA 180
QY 206 SDSGQDLFTNOLKEHPDTSVEFLSEONSNEVOEFAKAFAYTIEPOHRVDLQLYAPEA 265

DB 181 SPSDGDLLFTNQLKHPDTSVEFLQNSNEVOEFAFAFYIIEPQRDVLQLYAPEA 240

QY 266 FNYMDKFNE 274

DB 241 FNYMDKFNE 249

RESULT 8

US-08-082-849B-10

; Sequence 10, Application US/08082849B
; Patent No. 5677274

; GENERAL INFORMATION:

; APPLICANT: Leppla, Stephen H.

; APPLICANT: Klimpel, Kurt R.

; APPLICANT: Atora, Naveen

; APPLICANT: Singh, Yogendra

; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and

; NUMBER OF SEQUENCES: 35

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/082, 849B

; FILING DATE: 25-JUN-1993

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/021,601

; FILING DATE: 12-FEB-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Weber, Kenneth A.

; REGISTRATION NUMBER: 31,677

; REFERENCE/DOCKET NUMBER: 15280-161-1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0200

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 508 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-082-849B-10

Query Match 91.3%; Score 1282; DB 1; Length 508;

Best Local Similarity 100.0%; Pred. No. 7.5e-98;

Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGCHDVGAMHVEKEKNKDEERNKKTQSEHLKEIMKHIVKIEVGEEAVKKEAAE 85

DB 1 AGCHDVGAMHVEKEKNKDEERNKKTQSEHLKEIMKHIVKIEVGEEAVKKEAAE 60

QY 86 KLEKVPDYLEMYKAIGKIYIVDGDITKHISLEALSEDKKIKIDYKDALLHEHYV 145

DB 61 KLEKVPDYLEMYKAIGKIYIVDGDITKHISLEALSEDKKIKIDYKDALLHEHYV 120

QY 146 AKEGEPVAVIOSSEEDYVENTEKALNVYIEIGKILSRDILSKINOPYOKFLDVLNTIKNA 205

DB 121 AKEGEPVAVIOSSEEDYVENTEKALNVYIEIGKILSRDILSKINOPYOKFLDVLNTIKNA 180

QY 206 SPSDGDLLFTNQLKHPDTSVEFLQNSNEVOEFAFAFYIIEPQRDVLQLYAPEA 265

DB 181 SPSDGDLLFTNQLKHPDTSVEFLQNSNEVOEFAFAFYIIEPQRDVLQLYAPEA 240

QY 266 FNYMDKFNE 274

DB 241 FNYMDKFNE 249

RESULT 9

PCT-US94-01624-10

; Sequence 10, Application PC/TUS9401624
; GENERAL INFORMATION:

; APPLICANT: Leppla, Stephen H.

; APPLICANT: Klimpel, Kurt R.

; APPLICANT: Atora, Naveen

; APPLICANT: Singh, Yogendra

; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW

; STREET: Stewart Street Tower, 20th Floor, One Market

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94105

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/01624

; FILING DATE: June 25, 1993

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Weber, Kenneth A.

; REGISTRATION NUMBER: 31,677

; REFERENCE/DOCKET NUMBER: 15280-115

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 543-9600

; TELEFAX: (415) 543-5043

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 508 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; PCT-US94-01624-10

Query Match 91.3%; Score 1282; DB 5; Length 508;

Best Local Similarity 100.0%; Pred. No. 7.5e-98;

Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGCHDVGAMHVEKEKNKDEERNKKTQSEHLKEIMKHIVKIEVGEEAVKKEAAE 85

DB 1 AGCHDVGAMHVEKEKNKDEERNKKTQSEHLKEIMKHIVKIEVGEEAVKKEAAE 60

QY 86 KLEKVPDYLEMYKAIGKIYIVDGDITKHISLEALSEDKKIKIDYKDALLHEHYV 145

DB 61 KLEKVPDYLEMYKAIGKIYIVDGDITKHISLEALSEDKKIKIDYKDALLHEHYV 120

QY 146 AKEGEPVAVIOSSEEDYVENTEKALNVYIEIGKILSRDILSKINOPYOKFLDVLNTIKNA 205

DB 121 AKEGEPVAVIOSSEEDYVENTEKALNVYIEIGKILSRDILSKINOPYOKFLDVLNTIKNA 180

QY 206 SPSDGDLLFTNQLKHPDTSVEFLQNSNEVOEFAFAFYIIEPQRDVLQLYAPEA 265

DB 181 SPSDGDLLFTNQLKHPDTSVEFLQNSNEVOEFAFAFYIIEPQRDVLQLYAPEA 240

QY 266 FNYMDKFE 274
|
Db 241 FNYMDKFE 249

RESULT 10
US-08-021-601-2

; Sequence 2, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppia, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Nichols, Peter J.
; APPLICANT: Ariora, Naveen
; APPLICANT: Singh, Yogendra
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,601
; FILING DATE: 19930212
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-021-601-2

Query Match 91.3%; Score 1282; DB 1; Length 776;
Best Local Similarity 100.0%; Pred. No. 1.3e-97;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 AGHGDVGMVKKKKKKDKRDEERNKTOEHLKEIMKHIVKIVKGEAVKKEAAE 85
|
Db 1 AGHGDVGMVKKKKKKDKRDEERNKTOEHLKEIMKHIVKIVKGEAVKKEAAE 60
QY 86 KLEKVPDVLKEMVKAIGKIYVDGDTKHSISLEALSEDKKKIKDYGKDALHHEHYV 145
|
Db 61 KLEKVPDVLKEMVKAIGKIYVDGDTKHSISLEALSEDKKKIKDYGKDALHHEHYV 120
QY 146 AKEGYEPVLVIOSSSEDEVVENTERKALNYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 205
|
Db 121 AKEGYEPVLVIOSSSEDEVVENTERKALNYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 180
QY 206 SDSGDQLLFTTNQKHEPTPFSVEFLQNSNEVOEVAKAFAYITIEQHRDVLQLYAPEA 265
|
Db 181 SDSGDQLLFTTNQKHEPTPFSVEFLQNSNEVOEVAKAFAYITIEQHRDVLQLYAPEA 240
QY 266 FNYMDKFE 274
|
Db 241 FNYMDKFE 249

RESULT 11

US-08-082-849B-2
; Sequence 2, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
; APPLICANT: Leppia, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Ariora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/082,849B
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/021,601
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-161-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-082-849B-2

Query Match 91.3%; Score 1282; DB 1; Length 776;
Best Local Similarity 100.0%; Pred. No. 1.3e-97;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 AGHGDVGMVKKKKKKDKRDEERNKTOEHLKEIMKHIVKIVKGEAVKKEAAE 85
|
Db 1 AGHGDVGMVKKKKKKDKRDEERNKTOEHLKEIMKHIVKIVKGEAVKKEAAE 60
QY 86 KLEKVPDVLKEMVKAIGKIYVDGDTKHSISLEALSEDKKKIKDYGKDALHHEHYV 145
|
Db 61 KLEKVPDVLKEMVKAIGKIYVDGDTKHSISLEALSEDKKKIKDYGKDALHHEHYV 120
QY 146 AKEGYEPVLVIOSSSEDEVVENTERKALNYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 205
|
Db 121 AKEGYEPVLVIOSSSEDEVVENTERKALNYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 180
QY 206 SDSGDQLLFTTNQKHEPTPFSVEFLQNSNEVOEVAKAFAYITIEQHRDVLQLYAPEA 265
|
Db 181 SDSGDQLLFTTNQKHEPTPFSVEFLQNSNEVOEVAKAFAYITIEQHRDVLQLYAPEA 240
QY 266 FNYMDKFE 274
|
Db 241 FNYMDKFE 249


```

? TELEPHONE: (415) 617-8959
? TELEFAX: (415) 327-3331
? INFORMATION FOR SEQ ID NO: 4:
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 593 amino acids
?
? TYPE: AMINO ACID
?
? STRANDEDNESS: single
?
? TOPOLOGY: linear
?
? MOLECULE TYPE: peptide
US-07-961-522-4

```

Query Match	8.0%;	Score	113;	DB	1;	Length	593;
Best Local Similarity	24.3%;	Pred. No.	0.14;				
Matches	72;	Conservative	45;	Mismatches	91;	Indels	88;
						Gaps	19;

Qy	22	LVOGAGGHDVGMH-VVEKENKNOE---NKKRDEERNKTOEHLKIMKIVKIE---	VK 74
Db	17	LVASGCGSGFDLDPVRNPQAKAEKATTSVODEETFKKTRKEELDRIKMEPALGAYEIOILR	76
Qy	75	GEAEVKKRAAEKLEKEVSDYLEMYKALIGKITYIVDDIDRKHISLEALSDKKIKRINOY	134
Db	77	RNKAPKTTGEGKRNDR---VVELSEDKITRFLYOESVEIILPH--LDLPLN-KITTSNDVY-	128
Qy	135	KDALLHEH-----YVAKGGEYEVLIJOSSEYVENTEKALN-----	174
Db	129	HSMDKRLDKNDKDKVYRSGY---VYDOSFMEIRNDGCFHVFQOIGDYIYYL	179
Qy	175	-----EL--GKIILSR---DILEKINQIOPKFLDYLNTI---KNAS-----	209
Db	180	GVTSPSKELPKPKVIVSYKGTWIPFVNIN--LEKEIDGFDTSOGGANNASATISITETVNDHK	237
Qy	210	GODLLFTNQLK--EHPDSEVLEPQO-----NSNEQVEVPAKAFAYIE	251
Db	238	VGEKIGDNEVGVASHSEFAVDGPNKKILGSLYANGYINRRKADEYVTR---VSIE	290

RESULT 15
US-08-217-

US-08-217-438-4
; Sequence 4, Application US/08217438
; Patent No. 5521072

1 GENERAL INFORMATION:
 2 APPLICANT: Potter, Andrew A.
 3 APPLICANT: Gerlach, Gerald F.
 4 APPLICANT: Wilson, Philp J.
 5 APPLICANT: Rossi-Campos, Amalia
 6 TITLE OF INVENTION: ACTINOBACILLUS PLEUROPNEUMONIAE
 7 TITLE OF INVENTION: TRANSFERIN BINDING PROTEINS AND USES THEREOF
 8 NUMBER OF SEQUENCES: 11
 9 CORRESPONDENCE ADDRESS:

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/217,438
FILING DATE: 22-MAR-1994

ATTORNEY/AGENT INFORMATION:
NAME: Robins, Roberta L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9001-0015.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-617-8999
TELEFAX: 415-327-3231
TELEX:

```

: INFORMATION FOR SEQ ID NO: 4 :
: SEQUENCE CHARACTERISTICS:
: LENGTH: 593 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
US-08-217-438-4
```

Query Match	8.0%;	Score 113;	DB 1;	Length 593;
Best Local Similarity	24.3%;	Pred. No. 0.14;		
Matches	72;	Conservative	45;	Mismatches 91;
			Indels	88;
			Gaps	19;

```

00  17  LVACSGSGSFDELDVRPNQJAKAEKATTSYQDEBETKKTKEELDKLMEBALGETQILR  76
01  11  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
02  22  LVQSGAGGSDVGMH-VKREKENKDE---NKRKDEERNKTOEHLKIMKHIVKIE---VK  74
03  11  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
04  07  GEEAVKKAARLLEKVPDYLEMKALGKITYIVDGDITKHISLEALSDEKKIKINITYG  134
05  75  GEEAVKKAARLLEKVPDYLEMKALGKITYIVDGDITKHISLEALSDEKKIKINITYG  134
06  77  RNKAKTETTGKRNERR---VVELSEKTKITLYQESVEIIPH--LDELN-GKTTSDNDV-  128
07  135  KDALHHEH-----YYAAKEGYEPVLVIQSSSEYVENTEKALN-----VYY-  174
08  129  ---HMSDKRDLKRNOLKTVYRGY-----YDSSFMEIRNDGSHFVFKQIDGYIYL  179
09  175  -----ET--GKIISR---DILSKINQYOKFLDVLNTI---KNAS-----DSD  209
10  180  GVTSKEKLEPKRKVLSYKQTMQFVENIN--LEREIDGDTSGDGKNVSATSIETTVNRDHK  237
11  210  GODLFTFMOLK--EHPNDFSEVPELO-----NSNEQVEYAKKAPAYIIE  251
12  238  VGEKIGDNVEVGVAAHSEFAVDFOKKLTGSLYNGKINIRNKAKOEVKTR---YSIE  290

```

Search completed: December 2, 2001, 13:49:25
Job time: 205 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:49:43 ; Search time 55.64 Seconds
(without alignments)
225.680 Million cell updates/sec

Title: US-09-747-521-4_COPY_178_735
Perfect score: 2871
Sequence: 1 DNIGLPELKQKSSNSRRKRS.....LYISNPKNKVAVTKENT 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/6C.COMB.pep:*
6: /cgn2_6/prodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2758	96.1	735	1	US-08-021-601-4 Sequence 4, Appl1
2	2758	96.1	735	1	US-08-082-849B-4 Sequence 4, Appl1
3	2758	96.1	735	5	PCT-US94-01624-4 Sequence 4, Appl1
4	2758	96.1	903	1	US-08-021-601-12 Sequence 12, Appl1
5	2758	96.1	903	1	US-08-082-849B-12 Sequence 12, Appl1
6	2758	96.1	903	5	PCT-US94-01624-12 Sequence 12, Appl1
7	2581	89.9	719	1	US-08-082-849B-31 Sequence 31, Appl1
8	2581	89.9	719	5	PCT-US94-01624-31 Sequence 31, Appl1
9	591	20.6	881	4	US-08-960-780-32 Sequence 32, Appl1
10	591	20.6	881	4	US-09-073-898-32 Sequence 32, Appl1
11	589	20.5	884	1	US-08-471-033-5 Sequence 5, Appl1
12	589	20.5	884	2	US-08-471-044-5 Sequence 5, Appl1
13	589	20.5	884	2	US-08-463-483A-5 Sequence 5, Appl1
14	589	20.5	884	2	US-08-471-046A-5 Sequence 5, Appl1
15	589	20.5	884	2	US-08-470-566B-5 Sequence 5, Appl1
16	589	20.5	884	3	US-08-469-334-5 Sequence 5, Appl1
17	589	20.5	884	3	US-09-300-529-5 Sequence 5, Appl1
18	589	20.5	1346	1	US-08-471-033-23 Sequence 23, Appl1
19	589	20.5	1346	2	US-08-471-044-23 Sequence 23, Appl1
20	589	20.5	1346	2	US-08-463-483A-23 Sequence 23, Appl1
21	589	20.5	1346	2	US-08-471-046A-23 Sequence 23, Appl1
22	589	20.5	1346	2	US-08-470-566B-23 Sequence 23, Appl1
23	589	20.5	1346	2	US-08-469-334-23 Sequence 23, Appl1
24	589	20.5	1346	3	US-09-300-529-23 Sequence 23, Appl1
25	588	20.5	852	1	US-08-471-033-36 Sequence 36, Appl1
26	588	20.5	852	2	US-08-471-044-36 Sequence 36, Appl1
27	588	20.5	852	2	US-08-463-483A-36 Sequence 36, Appl1

28	588	20.5	852	2	US-08-471-046A-36	Sequence 36, Appl1
29	588	20.5	852	2	US-08-470-566B-36	Sequence 36, Appl1
30	588	20.5	852	2	US-08-469-334-36	Sequence 36, Appl1
31	588	20.5	852	3	US-09-300-529-36	Sequence 36, Appl1
32	588	20.5	1338	1	US-08-471-033-50	Sequence 50, Appl1
33	588	20.5	1338	2	US-08-471-044-50	Sequence 50, Appl1
34	588	20.5	1338	2	US-08-463-483A-50	Sequence 50, Appl1
35	588	20.5	1338	2	US-08-470-566B-50	Sequence 50, Appl1
36	588	20.5	1338	2	US-08-470-566B-50	Sequence 50, Appl1
37	588	20.5	1338	2	US-08-469-334-50	Sequence 50, Appl1
38	588	20.5	1338	3	US-09-300-529-50	Sequence 50, Appl1
39	582.5	20.3	784	4	US-09-371-913A-7	Sequence 7, Appl1
40	580	20.2	667	1	US-08-471-033-7	Sequence 7, Appl1
41	580	20.2	667	2	US-08-471-044-7	Sequence 7, Appl1
42	580	20.2	667	2	US-08-463-483A-7	Sequence 7, Appl1
43	580	20.2	667	2	US-08-471-046A-7	Sequence 7, Appl1
44	580	20.2	667	2	US-08-470-566B-7	Sequence 7, Appl1
45	580	20.2	667	2	US-08-469-334-7	Sequence 7, Appl1

ALIGNMENTS

RESULT 1
US-08-021-601-4
Sequence 4, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Kilmpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Anora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414,057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-4

Query Match 96.1%; Score 2758; DB 1; Length 735;
Best Local Similarity 96.8%; Pred. No. 3,7e-213;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;
QY 1 DNIGLPELKQKSSNSRRKRSAGPTVPDDNDGICPDSLEVEGTYVKKKRTFLSPWIS 60

```
Db 149 DNLQJPELKOKSSNRKRRKSTASGPTVPDRDNDGIPDSLEVEGYTVYDVKNKRTFLSPWIS 208
Qy 61 NIHEKKGILTKYKSSPEKSTASDPYSDEPEKVTGRIDKNVSPARHPVLAAYPIVHVDMEN 120
Db 209 NIHEKKGILTKYKSSPEKSTASDPYSDEPEKVTGRIDKNVSPARHPVLAAYPIVHVDMEN 268
Qy 121 ILSKNEOSTONTSEFTTISKNTSTRTHTSEVHGNAEVANTSTSTHTSEVHGNAE 180
Db 269 ILSKNEOSTONTSEFTTISKNTSTRTHTSEVHGNAEVANTSTSTHTSEVHGNAE 328
Qy 181 VHAVIDHSLSLAGERETAETGTLTADTARLANIRVYNTGTAPLYNVLPPTSLVLGN 240
Db 329 SSTVAIDHSLSLAGERETAETGTLTADTARLANIRVYNTGTAPLYNVLPPTSLVLGN 388
Qy 241 QTLATIKAKENQSOILAPNNYPSKNIAPIALNODDFSSPTIMNVOFLEKTKQL 300
Db 389 QTLATIKAKENQSOILAPNNYPSKNIAPIALNODDFSSPTIMNVOFLEKTKQL 448
Qy 301 RLDTQVYGNATYVNFENGRAVDGSMNSEVLPQIOETTARIIFNGKDLNVERRIA 360
Db 449 RLDTQVYGNATYVNFENGRAVDGSMNSEVLPQIOETTARIIFNGKDLNVERRIA 508
Qy 361 NPSDPLETTKPDMTLKEALKIAFGNEPNGNLQYOGKDTTEPFNFDOOTSONIKNO 420
Db 509 NPSDPLETTKPDMTLKEALKIAFGNEPNGNLQYOGKDTTEPFNFDOOTSONIKNO 568
Qy 421 LNATNITYVLDKIKILNAKNNILIRDKRFHYDRNNIYAVGADSEVVEAAREVINSSTEG 480
Db 569 LNATNITYVLDKIKILNAKNNILIRDKRFHYDRNNIYAVGADSEVVEAAREVINSSTEG 628
Qy 481 LNIDDKIRKILSGYIVETEDTGLKEVINDRYDMLNISLRDGTFTIDFKKYNKPL 540
Db 629 LNIDDKIRKILSGYIVETEDTGLKEVINDRYDMLNISLRDGTFTIDFKKYNKPL 688
Qy 541 ISNPYKVVAVYAKENT 558
Db 689 ISNPYKVVAVYAKENT 706

RESULT 2
US-08-082-849B-4
: Sequence 4, Application US/08082849B
: Patent No. 5677274
: GENERAL INFORMATION:
: APPLICANT: Leppla, Stephen H.
: APPLICANT: Klimpel, Kurt R.
: APPLICANT: Arota, Naveen
: APPLICANT: Singh, Yogendra
: APPLICANT: Nichols, Peter J.
: TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
: NUMBER OF SEQUENCES: 35
: CORRESPONDENCE ADDRESS:
: ADDRESS: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/082,849B
: FILING DATE: 25-JUN-1993
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/021,601
: FILING DATE: 12-FEB-1993
: ATTORNEY/AGENT INFORMATION:
```

```
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-4

Query Match
Best Local Similarity 96.1%; Score 2758; DB 1: Length 735;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

Db 1 DNLQJPELKOKSSNRKRRKSTASGPTVPDRDNDGIPDSLEVEGYTVYDVKNKRTFLSPWIS 60
Db 149 DNLQJPELKOKSSNRKRRKSTASGPTVPDRDNDGIPDSLEVEGYTVYDVKNKRTFLSPWIS 208
Qy 61 NIHEKKGILTKYKSSPEKSTASDPYSDEPEKVTGRIDKNVSPARHPVLAAYPIVHVDMEN 120
Db 209 NIHEKKGILTKYKSSPEKSTASDPYSDEPEKVTGRIDKNVSPARHPVLAAYPIVHVDMEN 268
Qy 121 ILSKNEOSTONTSEFTTISKNTSTRTHTSEVHGNAEVANTSTSTHTSEVHGNAE 180
Db 269 ILSKNEOSTONTSEFTTISKNTSTRTHTSEVHGNAEVANTSTSTHTSEVHGNAE 328
Qy 181 VHAVIDHSLSLAGERETAETGTLTADTARLANIRVYNTGTAPLYNVLPPTSLVLGN 240
Db 329 SSTVAIDHSLSLAGERETAETGTLTADTARLANIRVYNTGTAPLYNVLPPTSLVLGN 388
Qy 241 QTLATIKAKENQSOILAPNNYPSKNIAPIALNODDFSSPTIMNVOFLEKTKQL 300
Db 389 QTLATIKAKENQSOILAPNNYPSKNIAPIALNODDFSSPTIMNVOFLEKTKQL 448
Qy 481 LNIDDKIRKILSGYIVETEDTGLKEVINDRYDMLNISLRDGTFTIDFKKYNKPL 540
Db 629 LNIDDKIRKILSGYIVETEDTGLKEVINDRYDMLNISLRDGTFTIDFKKYNKPL 688
Qy 541 ISNPYKVVAVYAKENT 558
Db 689 ISNPYKVVAVYAKENT 706

RESULT 3
PCT-US94-01624-4
: Sequence 4, Application PC/TUS9401624
: GENERAL INFORMATION:
: APPLICANT: Leppla, Stephen H.
: APPLICANT: Klimpel, Kurt R.
: APPLICANT: Arota, Naveen
: APPLICANT: Singh, Yogendra
: APPLICANT: Nichols, Peter J.
: TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
```


ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
STREET: Steuart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-4

Query Match 96.1%; Score 2758; DB 5; Length 735;
Best Local Similarity 96.8%; Pred. No. 3.7e-213;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 DNQLPELKOKSSNRKRSTASAPYVDRDNDGIPDSLEVEGYTVVKKRRTFLSPWIS 60
DB 149 DNQLPELKOKSSNRKRSTASAPYVDRDNDGIPDSLEVEGYTVVKKRRTFLSPWIS 208
QY 61 NHEKKGLTKYKSSPEKWSSTASDPYSDFEKVTGRIDKNVSEPARHPLVAAPIYHVDMEN 120
DB 209 NHEKKGLTKYKSSPEKWSSTASDPYSDFEKVTGRIDKNVSEPARHPLVAAPIYHVDMEN 268
QY 121 IILSKNEDOSTQNTDSETRTISKNTSRTHTSEVHGNAEYHANTSTSRHTSEVHGNAE 180
DB 269 IILSKNEDOSTQNTDSETRTISKNTSRTHTSEVHGNAEYHANTSTSRHTSEVHGNAE 328
QY 181 VHAVAIDHSLSLAGERTWAETMGNTADTARLNANIRYVNTGTAPIYVNLPTTSLVLGKN 240
DB 329 SSTVAIDHSLSLAGERTWAETMGNTADTARLNANIRYVNTGTAPIYVNLPTTSLVLGKN 388
QY 241 QTLATIRAKENQLSQILAPNNYPSKNLAPIALNAODFSSPTITMNYNOFLELEKTOL 300
DB 389 QTLATIRAKENQLSQILAPNNYPSKNLAPIALNAODFSSPTITMNYNOFLELEKTOL 448
QY 301 RLDTDOYVGNIAIYVNPENGVRVDTGSMNSEVLPQIOETTARITFNKDLNLVRRILAAV 360
DB 449 RLDTDOYVGNIAIYVNPENGVRVDTGSMNSEVLPQIOETTARITFNKDLNLVRRILAAV 508
QY 361 NPSDPLETTKPDMTLKEALITAFGNPNPNCNLQOGKDIETFDENFPFOOTSONKKNOLAE 420
DB 509 NPSDPLETTKPDMTLKEALITAFGNPNPNCNLQOGKDIETFDENFPFOOTSONKKNOLAE 568
QY 421 LNAFTNIYVLDKIRLAKNKNILIRDKRFHYDRNNIAVGADESYYKAEAREVINSSTGLL 480
DB 569 LNAFTNIYVLDKIRLAKNKNILIRDKRFHYDRNNIAVGADESYYKAEAREVINSSTGLL 628
QY 481 LNTDKDKRKLISGYIIEPTGELKEVINDRYDMLNLTSSLRQDCKTIDFKKYNDKPLLY 540
DB 629 LNTDKDKRKLISGYIIEPTGELKEVINDRYDMLNLTSSLRQDCKTIDFKKYNDKPLLY 688
QY 541 ISNPNYKVNYAVATKENT 558

DB 689 ISNPNYKVNYAVATKENT 706

RESULT 4
US-08-021-601-12
Sequence 12, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Arcora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spralt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-12

Query Match 96.1%; Score 2758; DB 1; Length 903;
Best Local Similarity 96.8%; Pred. No. 5e-213;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 DNQLPELKOKSSNRKRSTASAPYVDRDNDGIPDSLEVEGYTVVKKRRTFLSPWIS 60
DB 149 DNQLPELKOKSSNRKRSTASAPYVDRDNDGIPDSLEVEGYTVVKKRRTFLSPWIS 208
QY 61 NHEKKGLTKYKSSPEKWSSTASDPYSDFEKVTGRIDKNVSEPARHPLVAAPIYHVDMEN 120
DB 209 NHEKKGLTKYKSSPEKWSSTASDPYSDFEKVTGRIDKNVSEPARHPLVAAPIYHVDMEN 268
QY 121 IILSKNEDOSTQNTDSETRTISKNTSRTHTSEVHGNAEYHANTSTSRHTSEVHGNAE 180
DB 269 IILSKNEDOSTQNTDSETRTISKNTSRTHTSEVHGNAEYHANTSTSRHTSEVHGNAE 328
QY 181 VHAVAIDHSLSLAGERTWAETMGNTADTARLNANIRYVNTGTAPIYVNLPTTSLVLGKN 240
DB 329 SSTVAIDHSLSLAGERTWAETMGNTADTARLNANIRYVNTGTAPIYVNLPTTSLVLGKN 388
QY 241 QTLATIRAKENQLSQILAPNNYPSKNLAPIALNAODFSSPTITMNYNOFLELEKTOL 300
DB 389 QTLATIRAKENQLSQILAPNNYPSKNLAPIALNAODFSSPTITMNYNOFLELEKTOL 448

QY 301 RLDTQVYGNATYVNFENGRAVRVDGSMSEVLPQIOETTARIIFNGKDLNVERIAAV 360
DB 449 RLDTQVYGNATYVNFENGRAVRVDGSMSEVLPQIOETTARIIFNGKDLNVERIAAV 508
QY 361 NPSDPLETTKPDMLKEALKIAFGNEPENGNIQYOGKDIETEDFENFDOTSONINQLA 420
DB 509 NPSDPLETTKPDMLKEALKIAFGNEPENGNIQYOGKDIETEDFENFDOTSONINQLA 568
QY 421 LNATNIYVLDKIKLNAMNIIIRDKRFHYDRNNIAGADESVKAEHREVINSSTEGIL 480
DB 569 LNATNIYVLDKIKLNAMNIIIRDKRFHYDRNNIAGADESVKAEHREVINSSTEGIL 628
QY 481 LNIDDKIRIISGYIVEIEDTEGLKEVINDRYDMLNISLRDGGTFIDFKKYNKPLX 540
DB 629 LNIDDKIRIISGYIVEIEDTEGLKEVINDRYDMLNISLRDGGTFIDFKKYNKPLX 688
QY 541 ISNPNKYVNVAVTKENT 558
DB 689 ISNPNKYVNVAVTKENT 706

RESULT 5
US-08-082-849B-12
Sequence 12, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arota, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082.849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-12

Query Match 96.1%; Score 2758; DB 1; Length 903;
Best Local Similarity 96.8%; Pred. No. 5e-213;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 DNLQPLPELKOKSSNRKRKRSTASGPTVPDRDNDGIPDSLEVEGYVYDVKNKRTFLSPMJS 60
DB 149 DNLQPLPELKOKSSNRKRKRSTASGPTVPDRDNDGIPDSLEVEGYVYDVKNKRTFLSPMJS 208
QY 61 NHEKKGLTKYKSSPEKKSTASDPYSDFEKYGRIDKNVSPARPLVAAPPIVHVDEN 120
DB 209 NHEKKGLTKYKSSPEKKSTASDPYSDFEKYGRIDKNVSPARPLVAAPPIVHVDEN 268
QY 121 ILSKNEQOSTONTDSETRTISKNTSTSRHTSEVHGAENAETHSTSRHTSEVHGA 180
DB 269 ILSKNEQOSTONTDSETRTISKNTSTSRHTSEVHGAENAETHSTSRHTSEVHGA 328
QY 181 VHAVALDHSLSIAGERTAEFTGLTADTARLANIRIVNTGTAPIVYVLPPTSVLCKN 240
DB 329 SSTVAIDHSLSIAGERTAEFTGLTADTARLANIRIVNTGTAPIVYVLPPTSVLCKN 388
QY 241 OTLATIKAKENOLSOILAPNNYPSKNLAPIALNODDESSPTIMNTNOFLEKTKOL 300
DB 389 OTLATIKAKENOLSOILAPNNYPSKNLAPIALNODDESSPTIMNTNOFLEKTKOL 448
QY 301 RLDTQVYGNATYVNFENGRAVRVDGSMSEVLPQIOETTARIIFNGKDLNVERIAAV 360
DB 449 RLDTQVYGNATYVNFENGRAVRVDGSMSEVLPQIOETTARIIFNGKDLNVERIAAV 508
QY 361 NPSDPLETTKPDMLKEALKIAFGNEPENGNIQYOGKDIETEDFENFDOTSONINQLA 420
DB 509 NPSDPLETTKPDMLKEALKIAFGNEPENGNIQYOGKDIETEDFENFDOTSONINQLA 568
QY 421 LNATNIYVLDKIKLNAMNIIIRDKRFHYDRNNIAGADESVKAEHREVINSSTEGIL 480
DB 569 LNATNIYVLDKIKLNAMNIIIRDKRFHYDRNNIAGADESVKAEHREVINSSTEGIL 628
QY 481 LNIDDKIRIISGYIVEIEDTEGLKEVINDRYDMLNISLRDGGTFIDFKKYNKPLX 540
DB 629 LNIDDKIRIISGYIVEIEDTEGLKEVINDRYDMLNISLRDGGTFIDFKKYNKPLX 688
QY 541 ISNPNKYVNVAVTKENT 558
DB 689 ISNPNKYVNVAVTKENT 706

RESULT 6
PCT-US94-01624-12
Sequence 12, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arota, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Stewart Street Tower, 20th Floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677

REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ. ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-12

Query Match 96.1%; Score 2758; DB 5; Length 903;
Best Local Similarity 96.8%; Pred. No. 5e-213;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 DNLQPELKOKSSNSRRKSTASGPTVPDRDNDGIPDSLEVEGYTVVKKRFTLSFWIS 60
DB 149 DNLQPELKOKSSNSRRKSTASGPTVPDRDNDGIPDSLEVEGYTVVKKRFTLSFWIS 208
QY 61 NIEKKGLTRYKSSPEKMSASDPYSDPEKYTGRIKNSPEARHPVLAAYPIVHVDMEN 120
DB 209 NIEKKGLTRYKSSPEKMSASDPYSDPEKYTGRIKNSPEARHPVLAAYPIVHVDMEN 268
QY 121 IILSKNEDOSTQNTDSETRTISKNTSRTHTSEVHGNAEVAHNTSRTHTSEVHGNAE 180
DB 269 IILSKNEDOSTQNTDSETRTISKNTSRTHTSEVHGNAEVAHNTSRTHTSEVHGNAE 328
QY 181 VHAVALDHSLSLAGEPTMAETMGLNTADTARLANINRYVTGTAPIVNLPPTSILVGN 240
DB 329 SSTVALDHSLSLAGEPTMAETMGLNTADTARLANINRYVTGTAPIVNLPPTSILVGN 388
QY 241 QTLATIKAKENQSLIAPNNYPSKNLADIALNAODFSSTPTITMNYNOFLELEKTKOL 300
DB 389 QTLATIKAKENQSLIAPNNYPSKNLADIALNAODFSSTPTITMNYNOFLELEKTKOL 448
QY 301 RLTDQVYGNATYVNFENGVRVDTGSMNSEVLPQIOETTARILFNKDLNLVERRIAAY 360
DB 449 RLTDQVYGNATYVNFENGVRVDTGSMNSEVLPQIOETTARILFNKDLNLVERRIAAY 508
QY 361 NPSDPLETTPDMTLKALKIAFGFNPNGNLQYOGKDITEFDNFPOQTSQNKKNOLAE 420
DB 509 NPSDPLETTPDMTLKALKIAFGFNPNGNLQYOGKDITEFDNFPOQTSQNKKNOLAE 568
QY 421 LNAATNITYVLDKIKLNKKNMILIRDKRFHYDRNNIAVGADESYYKEAAREVINSSTEGLL 480
DB 569 LNAATNITYVLDKIKLNKKNMILIRDKRFHYDRNNIAVGADESYYKEAAREVINSSTEGLL 628
QY 481 LNLDKDIRKILSGYIIEDETEGLEKVINRYDMLNLSIRROGKTFIDFKKYNDKPLV 540
DB 629 LNLDKDIRKILSGYIIEDETEGLEKVINRYDMLNLSIRROGKTFIDFKKYNDKPLV 688
QY 541 ISNPNKVVNYAATKENT 558
DB 689 ISNPNKVVNYAATKENT 706

RESULT 7
US-08-082-849B-31
Sequence 31, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Kilmpel, Kurt R.
APPLICANT: Airoa, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Fusion Proteins and
TITLE OF INVENTION: Related Methods
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET INFORMATION:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ. ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-31

Query Match 89.9%; Score 2581; DB 1; Length 719;
Best Local Similarity 90.9%; Pred. No. 5.7e-199;
Matches 511; Conservative 6; Mismatches 21; Indels 24; Gaps 2;

QY 1 DNLQPELKOKSSNS---KKRSTASGPTVPDRDNDGIPDSLEVEGYTVVKKRFTLS 56
DB 149 DNLQPELKOKSSNSMTATIMQGNFLOGPTVPDRDNDGIPDSLEVEGYTVVKKRFTLS 208
QY 57 PMISNIEKKGLTRYKSSPEKMSASDPYSDPEKYTGRIKNSPEARHPVLAAYPIVHV 116
DB 209 PMISNIEKKGLTRYKSSPEKMSASDPYSDPEKYTGRIKNSPEARHPVLAAYPIVHV 268
QY 117 DMENIILSKNEDOSTQNTDSETRTISKNTSRTHTSEVHGNAEVAHNTSRTHTSEVH 176
DB 269 DMENIILSKNEDOSTQNTDSETRTISKNTSRTHTSEVHGNAEVAHNTSRTHTSEVH 328
QY 177 GNAEVAHVAIDHSLSLAGEPTMAETMGLNTADTARLANINRYVTGTAPIVNLPPTSILV 236
DB 329 SNSNSSTVALDHSLSLAGEPTMAETMGLNTADTARLANINRYVTGTAPIVNLPPTSILV 388
QY 237 LGRNQTATIKAKENQSLIAPNNYPSKNLADIALNAODFSSTPTITMNYNOFLELEK 296
DB 389 LGRNQTATIKAKENQSLIAPNNYPSKNLADIALNAODFSSTPTITMNYNOFLELEK 439
QY 297 TKQLRLTDQVYGNATYVNFENGVRVDTGSMNSEVLPQIOETTARILFNKDLNLVERR 356
DB 440 -----YGNATYVNFENGVRVDTGSMNSEVLPQIOETTARILFNKDLNLVERR 488
QY 357 IAAVNPSPLETTTPDMTLKALKIAFGFNPNGNLQYOGKDITEFDNFPOQTSQNKKN 416
DB 489 IAAVNPSPLETTTPDMTLKALKIAFGFNPNGNLQYOGKDITEFDNFPOQTSQNKKN 548
QY 417 QLAELNATNITYVLDKIKLNKKNMILIRDKRFHYDRNNIAVGADESYYKEAAREVINSST 476
DB 549 QLAELNATNITYVLDKIKLNKKNMILIRDKRFHYDRNNIAVGADESYYKEAAREVINSST 608
QY 477 EGLLNLNDKDIRKILSGYIIEDETEGLEKVINRYDMLNLSIRROGKTFIDFKKYNDK 536
DB 609 EGLLNLNDKDIRKILSGYIIEDETEGLEKVINRYDMLNLSIRROGKTFIDFKKYNDK 668

QY 537 LPLYSNPNKYKVVAVTKENT 558
| | | | |
Db 669 LPLYSNPNKYKVVAVTKENT 690

RESULT 8

PCT-US94-01624-31
; Sequence 31, Application PC/TUS9401624
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Arota, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
; STREET: Stewart Street Tower, 20th Floor, One Market
; STREET: Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01624
; FILING DATE: June 25, 1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-115
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-01624-31

Query Match 89.9%; Score 2581; DB 5; Length 719;
Best Local Similarity 90.9%; Pred. No. 5.7e-199;
Matches 511; Conservative 6; Mismatches 21; Indels 24; Gaps 2;

QY 1 DNLQPELKQSSNS---RKKRSTSAQPTVPDRNDGIPDSLEVEGTVVYKKNRTPLS 56
| | | | |
Db 149 DNLQPELKQSSNTATIMQGNFLQGPYPVDRNDGIPDSLEVEGTVVYKKNRTPLS 208
QY 57 PWISNIHKKGLTKYKSSPEKMTASDPYDEKVTGRIDKVNSEARHPLVAAPYIVYV 116
| | | | |
Db 209 PWISNIHKKGLTKYKSSPEKMTASDPYDEKVTGRIDKVNSEARHPLVAAPYIVYV 268
QY 117 DMENIILSKNEDQSTQNTDSETRTTSKNTSTSRHTTSEVHGNAEYHANTSTSRHTTSEVH 176
| | | | |
Db 269 DMENIILSKNEDQSTQNTDSETRTTSKNTSTSRHTTSEVHGNAEYHANTSTSRHTTSEVH 328
QY 177 GNAEYHAVAIDHSLSLAERTMAETMGUNTADTARLANIRVYNGTAPIYVLTPTSLV 236
| | | | |
Db 329 SNSNSTVAIDHSLSLAERTMAETMGUNTADTARLANIRVYNGTAPIYVLTPTSLV 388
QY 237 LGRNOTLATIRAKENQLSOILAPNNYVPSKNLAPIALNADODESSPTTMVNOGLELEK 296
| | | | |
Db 389 LGRNOTLATIRAKENQLSOILAPNNYVPSKNLAPIALNADODESSPTTMVNOGLELEK 439

QY 297 TKQLRLTDQVYGNATYVNFENGRRVVDTSNMSEVLPQIQETTARIIFNGKDLNIVER 356
| | | | |
Db 440 -----YGNATYVNFENGRRVVDTSNMSEVLPQIQETTARIIFNGKDLNIVER 488
QY 357 IAAVNSDPLETTKPDMTLKEALKIAFGNEPNGLQYQKDIETEDPNDQOSTONIKN 416
| | | | |
Db 489 IAAVNSDPLETTKPDMTLKEALKIAFGNEPNGLQYQKDIETEDPNDQOSTONIKN 548
QY 417 QLAELNATVTVLQKIKINAKMNLIRKRPYDRNNIAGVADSVYEAHREYINNST 476
| | | | |
Db 549 QLAELNATVTVLQKIKINAKMNLIRKRPYDRNNIAGVADSVYEAHREYINNST 608
QY 477 ECLLNIDKDIRKILSGYVEIEDETEGLKEVINDRYDMLNLSLRDQKTFIDFKKYNK 536
| | | | |
Db 609 ECLLNIDKDIRKILSGYVEIEDETEGLKEVINDRYDMLNLSLRDQKTFIDFKKYNK 668

RESULT 9

US-08-960-780-32
; Sequence 32, Application US/08960780
; Patent No. 6204435
; GENERAL INFORMATION:
; APPLICANT: Felleison, Jerald S.
; APPLICANT: Schepf, H. Ernest
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Stockhoff, Brian A.
; APPLICANT: Schmeltz, James
; APPLICANT: Loewer, David
; APPLICANT: Dullum, Charles Joseph
; APPLICANT: Muller-cohn, Judy
; APPLICANT: Stamp, Lisa
; TITLE OF INVENTION: No. 6204435el Pesticidal Toxins and Nucleotide
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,780
; FILING DATE: 30-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA-708
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 881 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: 177C8
US-08-960-780-32

Query Match 20.6%; Score 591; DB 4; Length 881;
Best Local Similarity 29.9%; Pred. No. 4.3e-39;
Matches 180; Conservative 97; Mismatches 218; Indels 108; Gaps 22;

QY 1 DNLQPELKOKSS-----NSRKRSTASGPTVDRNDGIPDSLEVBGYVD 47
DB 186 DELRNPEFNKESOEFLAKPSKINLFTQMKMREIDED---TDTGDSIPDLMEENGTYI- 241
QY 48 VKNKRTFLSPWISNIHEKGLTKYKSSPEKMWSTADPYSDPEKYTGRIKDVSPARHPL 107
DB 242 ---QNRIVAKWDDSL-ASKGTYKFSNPLSHYVGPYTDYKARLDLSNAKETNP 297
QY 108 VAAVPIVHVDMENILSKNEDOSTONTDSETRTISKNTSRTHTSEVHGNAEYHANTST 167
DB 298 VAAFPVSVMSEKVIILSPNENLS-----NSVESHSSTNW 331
QY 168 SRHTSEVHGNAEYH-----AVAIHSLSLAGERTWAEKMG---LNTADTARLANAN 215
DB 332 SYNTTEGASVEAGIGPKGISFGVSVNOHSEYVAQE--WGTSTGNTSOFTASAGYULAN 389
QY 216 IRYVNTGAPVINYVLTPTSLVGLKNOTLATIKAKENOLSOILAPNNYPSKNLAPIALNA 275
DB 390 VRIANNVGTGALIVDKPTTSYVL-NDITATITTAKSNTALNISGSESTPKKGONGIAITS 448
QY 276 QDDFSPTITMNYNOFLELEKTKQLRLDQVYGNIAITYNEENGRRVVDGSSNSEVLPQ 335
DB 449 MDDFNSHPTILNKQVONLNNKPMLETNQTDG---VYKIKDTHGNIVTGGEWNGYIQQ 505
QY 336 IOETTARIINGKDLNVEERIAAVNPSDPLETTKPDMLKEAKIAF--GFNEPNCMLQ 393
DB 506 IKAFTASIIYVDGE-RVAEKRVAAKDYENPEDKT-PSLTJLKDALKLSYPOEIKIEGLLY 563
QY 394 YQKGDITEFDF--NFDQOTSQINIKQLAEL-----NATNITYVDKIKLAKMMLIIRDK 446
DB 564 YKKNPIYESSVMYLDENTFAKEVTKQJLDPTGKFKDYSHLXYD---KLTPKMAVTYIK-L 618
QY 447 RFIYDRNNIVGADESVEAKHREVINSSTEG-----LLNID-----KDIR 488
DB 619 SILYDN---AESNDNSIGKWTNTNIVSGNNGKKQYSSNPDANLTLTNDQOEKLRKMRD 675
QY 489 KILSGYVELEDREGLKEVINDRYDMLNIS-SLRQDGTFFDEKKYNDKPLIYSNPYK 547
DB 676 YVLSLYKSKSEKNTQCEITIDGELYPIITTKYVNNKDKYKRLDIIAHNIK-----SNPISS 730
QY 548 VNV 550
DB 731 IHI 733

RESULT 10
US-09-073-898-32
Sequence 32, Application US/09073898
Patent No. 6242669
GENERAL INFORMATION:
APPLICANT: Feltelson, Gerald S.
APPLICANT: Schepf, H. Ernest
APPLICANT: Narva, Kenneth E.
APPLICANT: Stockhoff, Brian A.
APPLICANT: Schmelts, James
APPLICANT: Loewer, David
APPLICANT: Dullum, Charles Joseph
APPLICANT: Muller-Cohn, Judy
APPLICANT: Stamp, Lisa
APPLICANT: Morrill, George
APPLICANT: Finstad-Lee, Stacey
TITLE OF INVENTION: No. 6242669el Pesticidal Toxins and Nucleotide
TITLE OF INVENTION: Sequences Which Encode These Toxins
NUMBER OF SEQUENCES: 144
CORRESPONDENCE ADDRESS:

ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606-6669
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073,898
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,848
FILING DATE: 30-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/960,780
FILING DATE: 30-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sanders, Jay M.
REGISTRATION NUMBER: 39,355
REFERENCE/DOCKET NUMBER: MA-708C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 881 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: PS177C8
US-09-073-898-32

Query Match 20.6%; Score 591; DB 4; Length 881;
Best Local Similarity 29.9%; Pred. No. 4.3e-39;
Matches 180; Conservative 97; Mismatches 218; Indels 108; Gaps 22;

QY 1 DNLQPELKOKSS-----NSRKRSTASGPTVDRNDGIPDSLEVBGYVD 47
DB 186 DELRNPEFNKESOEFLAKPSKINLFTQMKMREIDED---TDTGDSIPDLMEENGTYI- 241
QY 48 VKNKRTFLSPWISNIHEKGLTKYKSSPEKMWSTADPYSDPEKYTGRIKDVSPARHPL 107
DB 242 ---QNRIVAKWDDSL-ASKGTYKFSNPLSHYVGPYTDYKARLDLSNAKETNP 297
QY 108 VAAVPIVHVDMENILSKNEDOSTONTDSETRTISKNTSRTHTSEVHGNAEYHANTST 167
DB 298 VAAFPVSVMSEKVIILSPNENLS-----NSVESHSSTNW 331
QY 168 SRHTSEVHGNAEYH-----AVAIHSLSLAGERTWAEKMG---LNTADTARLANAN 215
DB 332 SYNTTEGASVEAGIGPKGISFGVSVNOHSEYVAQE--WGTSTGNTSOFTASAGYULAN 389
QY 216 IRYVNTGAPVINYVLTPTSLVGLKNOTLATIKAKENOLSOILAPNNYPSKNLAPIALNA 275
DB 390 VRIANNVGTGALIVDKPTTSYVL-NDITATITTAKSNTALNISGSESTPKKGONGIAITS 448
QY 276 QDDFSPTITMNYNOFLELEKTKQLRLDQVYGNIAITYNEENGRRVVDGSSNSEVLPQ 335
DB 449 MDDFNSHPTILNKQVONLNNKPMLETNQTDG---VYKIKDTHGNIVTGGEWNGYIQQ 505
QY 336 IOETTARIINGKDLNVEERIAAVNPSDPLETTKPDMLKEAKIAF--GFNEPNCMLQ 393
DB 506 IKAFTASIIYVDGE-RVAEKRVAAKDYENPEDKT-PSLTJLKDALKLSYPOEIKIEGLLY 563
QY 394 YQKGDITEFDF--NFDQOTSQINIKQLAEL-----NATNITYVDKIKLAKMMLIIRDK 446

Db 564 YAKKPIYESVWTFYLDENTAKETKQKLDVSHLYDV-----KLTPrMANTIK-L 618
QY 447 RHHYRNNIAGADESVYKAEHREVIINSSTEG-----LLNID-----KDIR 488
Db 619 SLTYDN---AESNDNSIGKWTNTNIVSGNNGKKOYSSNPNPDLNTDADEKLNKND 675
QY 489 KILSGYIEIEETEGLEKYINDRYMLNIS-SLRDGGKTFIDFKKYNDKLPYISNPYK 547
Db 676 YIISLYMSEKNTQCEITIDGELIPIITTKTVNKNKDYKRLDIHANIK-----SNPISS 730
QY 548 VNV 550
Db 731 IHI 733

RESULT 11
US-08-471-033-5
; Sequence 5, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostlichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESS: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.308
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471.033
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-471-033-5

Query Match 20.5%; Score 589; DB 1; Length 884;
Best Local Similarity 30.2%; Pred. No. 6,3e-39;
Matches 180; Conservative 94; Mismatches 215; Indels 108; Gaps 22;

QY 1 DNLQPELKOKSS-----NSRKRSTASGPIVPRDNDGIPDSLEVGTYVD 47
Db 189 DELRNEEFENKKEQOEPLAKPSKINLFTQMKREIDED---TPTDDSDIPDLWEEGTYI- 244
QY 48 YNKRRTFLSPWISNIHEKKGLTKFYKSPPEKMWSTADPYSDPEKVTGRIDKNVSPARHPL 107
Db 245 ---QNRIVAKWDDSL-ASGTYTKFVSNPLESHTYVDPTDYKARDDLSNAKETFNPL 300
QY 108 VAAPIVHVDMENILSKNEDOSTONTDSEPTTKSTKNTSTSRTHSEVHGAENVAHNTST 167
Db 301 VAAFPVSNVSMKEVILSPENLS-----NSVESHSSTNW 334
QY 168 SRTHSEVHGAENVAH-----VAIDHSLSLAGERTWAEIMG-----LNTADTARLANN 215
Db 335 STYNTGASVEAGIGPKGISFGSVNYQHSETVAGE--WGISTGNTSOFNTASAGYLANN 392
QY 216 IRYVMTGAPRIYVLPPTSLVLGKNOTLATIKAKENQSLQILAPNNYPSKNLAPIALNA 275
Db 393 VAYNVNGCAIYDVAKPTTSFVL--NNDTIATITAKSNSTRALNISPESEYPKKONGIAITS 451
QY 276 ODEFSPTITMNYNOFLEKTKQLRLDTQYGNIAFYNEFNGRVAVDTSNMSEVLPO 335
Db 452 MDFFNSHPTTLNKKQVDNLNKKPMULETNTQDG---YKIKIDTGNITVGGEMNGVIOQ 508
QY 336 IOETARIITFNCKDLNIVERRIAAVNPSDPLETTKPDMTIKALKATIAF--GNEENGNIQ 393
Db 509 IAKTASITVDGE-RVAKERYAAKQYENPEDKT-PSLTKALKALSLPDEIKETEGELY 566
QY 394 YQKDIITEFD--NPDQOTSONIKNOAEL-----NATNIYVLDKIKLAKMANILIRDK 446
Db 567 YNKRPIYESVWTFYLDENTAKETKQKLDVSHLYDV-----KLTPrMANTIK-L 621
QY 447 RHHYRNNIAGADESVYKAEHREVIINSSTEG-----LLNID-----KDIR 488
Db 622 SLTYDN---AESNDNSIGKWTNTNIVSGNNGKKOYSSNPNPDLNTDADEKLNKND 678
QY 489 KILSGYIEIEETEGLEKYINDRYMLNIS-SLRDGGKTFIDFKKYNDKLPYISNP 544
Db 679 YIISLYMSEKNTQCEITIDGELIPIITTKTVNKNKDYKRLDIHANIK-----SNP 730
RESULT 12
US-08-471-044-5
; Sequence 5, Application US/08471044
; Patent No. 5840868
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostlichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESS: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.308

;;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,044
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-471-044-5

Query Match 20.5%; Score 589; DB 2; Length 884;
Best Local Similarity 30.2%; Pred. No. 6,3e-39;

Matches 180; Conservative 94; Mismatches 215; Indels 108; Gaps 22;

QY 1 DNOLPELKOKSS-----NSRKRRTSAGPTVPDRDNDGIPDSLEVEGYTVD 47
DB 189 DELRNPFPNKKESQEFLLAKSKINLTQKKMKREIDED---TDTGDSIPDLMEENGTTI- 244
QY 48 VKNKRRTFLSPWISNIHEKKGLTKYKSSPEKWSSTASDPYSDPEKYTGRIKDVNSPEARHPL 107
DB 245 ---QNRILAVKWDLSL-ASKGYTKFVSNPLSHVYGDYTYDEKAAARLDLSNAKETFNPL 300
QY 108 VAAFPYIVHDMENIILSKNDOSTONTSETRTISKNTSTSRHTSEVHGNAEYHANTST 167
DB 301 VAAFPYVSNVSMEXVILSPNENLS-----NSVESHSSTNW 334
QY 168 SRPHTSEVHGNAEYH-----AVADHSLSLAGERTMAETWG---LNTADTARILAN 215
DB 335 STYNTGASVEAGIGPKGISFGVSNVQHSSETVAQE--WGTSTGNSTQFNTASAGYLTAN 392
QY 216 IRYVNTGTAPIVNVLTPTSLVLGKNOTLATIKAKENOLSOILAPNNYPSKNLAPIALNA 275
DB 393 VRYNNTGTGALYDVKKPTTSFVL--NNDTIATITAKSNSTALNISGEEYPRKGGNGIATIS 451
QY 276 QDDFSSTPTMANOQLELEKTKQLRLDTQYGNLATYFENGRAVAVDTGSMSEVLPQ 335
DB 452 MDDFNHPITLANKQVDNLNNKPMLETTNOTG---VYKIKDTHGNIIVYGGENGVIYQ 508
QY 336 IOETRTITFNGKDLNVERIRIAAVNPDSLETTTKPDMTEKALKIAF--GFNPNPNLQ 393
DB 509 IKTAKTISIYDGE--RVAERKRVAKTDENPEDXT--PSLTIKDLKLSYPEIETIEBLLY 566
QY 394 YQGRDITEDF--NPDQTSQNIKNOLAEL-----NATNIYTVLDKIKLAKNNILIRDK 446
DB 567 YKRPPIYESSVMTYDENTAKETVKQANDTGTGFKVDYSHLYDV---KLPPKNAVITIK-L 621
QY 447 RFIYDRNNINAVGDESVEYKAEHREVINSSTEG-----LLINID-----KDIR 488
DB 622 SILYDN---AESNDNSIGKWTNTNIVSGCNGNKKOYSSNNPDANLLTLNTDAQEKLNKRD 678
QY 489 KILSGIYVEIEDPEGLKEVINDRYDMLNIS--SLRQDCKTFIDKKYNDKLPYISNP 544

DB 679 YVLSLVKMSKKNRQCELTITGEIETPITTKYVNAKKDKVNRDLITAHNIK-----SNP 730

;;
; RESULT 13
; US-08-463-483A-5
; Sequence 5, Application US/08463483A
; Patent No. 5849870
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5849870el Pesticidal proteins and strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,483A
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murtry
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8615
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-463-483A-5

Query Match 20.5%; Score 589; DB 2; Length 884;
Best Local Similarity 30.2%; Pred. No. 6,3e-39;

Matches 180; Conservative 94; Mismatches 215; Indels 108; Gaps 22;

QY 1 DNOLPELKOKSS-----NSRKRRTSAGPTVPDRDNDGIPDSLEVEGYTVD 47
DB 189 DELRNPFPNKKESQEFLLAKSKINLTQKKMKREIDED---TDTGDSIPDLMEENGTTI- 244
QY 48 VKNKRRTFLSPWISNIHEKKGLTKYKSSPEKWSSTASDPYSDPEKYTGRIKDVNSPEARHPL 107
DB 245 ---QNRILAVKWDLSL-ASKGYTKFVSNPLSHVYGDYTYDEKAAARLDLSNAKETFNPL 300

APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalin M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5872212el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5872212artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,566B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meligs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-566B-5

Query Match 20.5%; Score 589; DB 2; Length 884;
Best local Similarity 30.2%; Pred. No. 6, 3e-39;
Matches 180; Conservative 94; Mismatches 215; Indels 108; Gaps 22;

QY 1 DNLQLELKOKSS-----NSRKKRSTSGPTVPDRDNDGIPDSLEVEGYTD 47
DB 189 DELRNPENFKKESQEFLLAKPSKINLFTQKKREIDED---TDTGDSIPDLMEENGTYI- 244
QY 48 VKKKRFLSPWISNIEHKKLTYYKSSPEKMWSTASDPYSDPEKYTGRIIDKNVSPKARHPL 107
DB 245 ---QNRILAVWMDSL-ASKGYTKFVSNPLSHVGVDPYTDYEKAARDLDSNAKETFNPL 300
QY 108 VAAVPIYVDMENIILSKNEDOSTQNTDSETRTISKNTSTSRHTSEVHGNAEVAHANTST 167
DB 301 VAAFPVYVNMKEVILSPNENLS-----NSVESHSSTNW 334
QY 168 SRHTSEVHGNAEYH-----AVAIHSLSLAGERTMAETMG---LNTADTARLNNAN 215
DB 335 SYNTTEGASVEAIGPGIGISFVGSVNYQHSETVAQE--MCTSTGNTSQPNTASAGYLNAN 392

QY 216 IRVYNTGTADIVNVLPPTSLVLGKNOFLATIKAKENOLSOILAPNNYPSKNLAPIALNA 275
DB 393 VRVNNVGTGATLYDKPRTTSVL--NNDIATITAKSNSTALNISGESYPRKGGONGIATIS 451
QY 276 QDDSSPTITMANNQFLELEKTKQLRLDQVYGNIAITYFENGVRVVDYTGSMNSEVLPO 335
DB 452 MDDNSHPITLNNKQVDNLNNKPMLETNQTDG---VYKIKDTHGNIYVGGEWNGVIOQ 508
QY 336 IOETTAIINGKDLNIVERIRIAVNSDPLETTKPMTLKALKIAF--GFNPENGLQ 393
DB 509 IKAKTASIIIVDGE--RVAEKRVAAKDYENPEDKT--PSLTLKDALKLSYPOEIKETIGILY 566
QY 394 YOGKDIREPFD--NFDQOTSONIKOLAEL-----NATNIYTYVADTKLAKNNILRDK 446
DB 567 YKKNPIYESSVMYTLDENTAKYVKQLNDITGAKKDYSHLYDV---KLTPKMNVTIK-L 621
QY 447 RFHYDRNNIIVGADSVYKKAHREVINSSTEG-----LLNID-----KDIR 488
DB 622 SILYDN---AESNDNSIGKMTNTNIVSGNNGKKQYSSNPDAN/LTNTDAQEKLKNNRD 678
QY 489 KILSGYIVLEIDTEGLKEVINDRYDMLNIS--SLRQDKTFIDFKKYDKLPLYSNP 544
DB 679 YVLSLYMKSEKNTQCEITIDGELIPIYTKYVNVNKKDYKKRLDIAHNK-----SNP 730

Search completed: December 2, 2001, 13:50:01
Job time: 241 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:54:38 ; Search time 132.69 seconds
(without alignments)
842.204 Million cell updates/sec

Title: US-09-747-521-4
Perfect score: 3913
Sequence: 1 MKRRKVLPLMALSTLVSS.....TSTNGIKILIFSKGYEIG 764

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPTREMBL_17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mhc:*
8: sp_mammal:*
9: sp_organelle:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3797	97.0	764	2	09RQ02
2	3793	96.9	764	2	09KH69
3	3785	96.7	764	2	09F5R7
4	900	23.0	876	2	032739
5	898	22.9	876	2	09KH41
6	886	22.6	879	2	006498
7	883.5	22.6	875	2	046221
8	798	20.4	721	2	086171
9	235.5	6.0	204	2	09X377
10	205.5	5.3	4688	2	09PQ08
11	199.5	5.1	1387	5	09G276
12	197	5.0	2647	5	09U4X0
13	195.5	5.0	3254	5	09BK45
14	193.5	4.9	3130	5	09BK46
15	193.5	4.9	4919	2	09GH10
16	193	4.9	1072	2	09CF64
17	190	4.9	1302	2	049547
18	189	4.8	6713	2	099054
19	188.5	4.8	2771	5	026216

20	186	4.8	2178	2	046149	046149	clostridium
21	182.5	4.7	1193	2	045914	045914	clostridium
22	182	4.7	2747	5	09BJX9	09BJX9	plasmidium
23	180.5	4.6	149	2	09RM77	09RM77	clostridium
24	180	4.6	1051	2	049524	049524	mycoplasma
25	180	4.6	2269	5	026223	026223	plasmidium
26	179.5	4.6	962	2	049546	049546	mycoplasma
27	179.5	4.6	1365	2	049525	049525	mycoplasma
28	179	4.6	769	2	099V70	099V70	staphylococ
29	178.5	4.6	149	2	09RM78	09RM78	clostridium
30	178.5	4.6	1193	2	P71107	P71107	clostridium
31	177	4.5	1315	2	086488	086488	staphylococ
32	177	4.5	1837	3	074424	074424	schizosacch
33	177	4.5	3724	5	077320	077320	plasmidium
34	176.5	4.5	149	2	09RM80	09RM80	clostridium
35	176	4.5	1621	5	024984	024984	giardia lam
36	175.5	4.5	1516	5	096154	096154	plasmidium
37	175	4.5	1176	5	09F231	09F231	helicobacte
38	175	4.5	1939	5	025662	025662	plasmidium
39	174.5	4.5	149	2	09RM79	09RM79	helicobacte
40	174	4.4	2529	2	025579	025579	plasmidium
41	173.5	4.4	1344	2	049545	049545	mycoplasma
42	173	4.4	1230	2	025772	025772	helicobacte
43	171.5	4.4	821	3	008581	008581	saccharomyc
44	171.5	4.4	1098	5	09C275	09C275	plasmidium
45	170	4.3	1152	4	092603	092603	homo sapien

ALIGNMENTS

RESULT 1
ID 09RQ02; PRELIMINARY; PRT; 764 AA.
AC 09RQ02;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PX01-110 (PROTECTIVE ANTIGEN).
GN PNG.
OS Bacillus anthracis.
OC Plasmid virulence plasmid PX01, and Plasmid PX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPERNE; PLASMID=VIRULENCE PLASMID PX01;
RX MEDLINE=99445483; PubMed=10515943;
RA Okinaka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hill K.K.,
RA Keim P., Koehler T.M., Lamke G., Kumano S., Mahillon J., Manter D.,
RA Martinez Y., Rieke D., Svensson R., Jackson P.J.,
RT "Sequence and organization of px01, the large Bacillus anthracis
RT plasmid harboring the Anthrax toxin genes.";
RL J. Bacteriol. 181:6509-6515(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=33, 28, AND BA1035; PLASMID=PX01;
RX MEDLINE=99214082; PubMed=10197996;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.,
RT "Genetic diversity in the protective antigen gene of Bacillus
RT anthracis.";
RL J. Bacteriol. 181:2358-2362(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=33, 28, AND BA1035; PLASMID=PX01;
RX Price L.B., Hugh-Jones M., Jackson P.J., Keim P.,
RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF063404; AAD32414.1;
DR EMBL: AF063404; AAD32414.1;
DR EMBL: AF306781; AAG24449.1;
DR EMBL: AF306778; AAG24448.1;
DR EMBL: AF306779; AAG24447.1;
DR EMBL: AF306780; AAG24448.1;
DR EMBL: AF306780; AAG24448.1;

DR HSSP: P13423; 1ACC.
DR InterPro: IPR003896; Binary_toxB.
DR PRINTS: PR01391; BINARYTOXINB.
KW Plasmid.
SO SEQUENCE 764 AA; 85810 MW; 3AE1E8BF48FAA03F CRC64;

Query Match 97.0%; Score 3797; DB 2; Length 764;
Best Local Similarity 97.5%; Pred. No. 3.3e-208;
Matches 745; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

```
QY 1 MKKRVLLPMAALSTIIYSSGNGNEVIAEYKQENRLNLSSESSSGGLGYFSDLNQA 60
DB 1 MKKRVLLPMAALSTIIYSSGNGNEVIAEYKQENRLNLSSESSSGGLGYFSDLNQA 60
QY 61 PMVYTSSTTGDISPSSSELENIPSENQYFQSAIMSGFTIKVKSDEYTPATSDNHVTMW 120
DB 61 PMVYTSSTTGDISPSSSELENIPSENQYFQSAIMSGFTIKVKSDEYTPATSDNHVTMW 120
QY 121 DDQEVINKASNSNKTIRLEKGRLYQIKIOYORENPTKEGLDFKLWYTDSONKKEVISSDNL 180
DB 121 DDQEVINKASNSNKTIRLEKGRLYQIKIOYORENPTKEGLDFKLWYTDSONKKEVISSDNL 180
QY 181 QLPETKOKSSNRKRRTSAGPTVPDRNDGIPDSLEVEGYVDVKNKRTFLSPWISNH 240
DB 181 QLPETKOKSSNRKRRTSAGPTVPDRNDGIPDSLEVEGYVDVKNKRTFLSPWISNH 240
QY 241 EKKGLTKKSSPEKMTASDPYSDPEKTYGRIDKNVSEARHPPLVAAPYIVHDMENITL 300
DB 241 EKKGLTKKSSPEKMTASDPYSDPEKTYGRIDKNVSEARHPPLVAAPYIVHDMENITL 300
QY 301 SKNEDOSTQNTDSETRTISKNTSTSRHTSEVHGNAEYHANTSTSRHTSEVHGNAEYHA 360
DB 301 SKNEDOSTQNTDSETRTISKNTSTSRHTSEVHGNAEYHANTSTSRHTSEVHGNAEYHA 360
QY 361 VAIDHSLSLAGERTAEETMGNTADTARLANIRVNTGTAPYVAVLPTTSLVLCGNQTL 420
DB 361 VAIDHSLSLAGERTAEETMGNTADTARLANIRVNTGTAPYVAVLPTTSLVLCGNQTL 420
QY 421 ATIRAKENQSOIILAPNNYPSKNLAPIALNADDFSSPTITMANNQFLEKTKQLRLD 480
DB 421 ATIRAKENQSOIILAPNNYPSKNLAPIALNADDFSSPTITMANNQFLEKTKQLRLD 480
QY 481 TDQVYGNATYTNFENGRRVVDGTSNMSEVLPOIOETTAIIIFNGKDLNVERRIAANPS 540
DB 481 TDQVYGNATYTNFENGRRVVDGTSNMSEVLPOIOETTAIIIFNGKDLNVERRIAANPS 540
QY 541 DPLETTKPDMLTKEALKTAFGNPNNGMLQYOGKDTFEDFNFDOQTSONIKNOAELNA 600
DB 541 DPLETTKPDMLTKEALKTAFGNPNNGMLQYOGKDTFEDFNFDOQTSONIKNOAELNA 600
QY 601 TNYIVLADIKILNKAAMNLIIRDREHYDRNNIAVGADESUYKFAHREVIINSSTEGILLNI 660
DB 601 TNYIVLADIKILNKAAMNLIIRDREHYDRNNIAVGADESUYKFAHREVIINSSTEGILLNI 660
QY 661 DKDIRKILSGYIVIEIDTEGLKEVINDRYDMLNISLRQDGKTFIDFKYNDKLPYLYSN 720
DB 661 DKDIRKILSGYIVIEIDTEGLKEVINDRYDMLNISLRQDGKTFIDFKYNDKLPYLYSN 720
QY 721 PNYKVNVAATKENTTIINPSENGDSTNGIKKILFSSKKGYEIG 764
DB 721 PNYKVNVAATKENTTIINPSENGDSTNGIKKILFSSKKGYEIG 764
```

RESULT 2

Q9KH69 PRELIMINARY; PRT; 764 AA.
AC Q9KH69;
DT 01-OCT-2000 (TReMBLrel. 15; Created)
DT 01-OCT-2000 (TReMBLrel. 15; Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17; Last annotation update)
DE PROTECTIVE ANTIGEN.
GN PAGE OR PAG.

OS Bacillus anthracis.
OG Plasmid pX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;

```
RA [1]
RA SEQUENCE FROM N.A.
RA STRAIN-V770-NP1-R, ATCC14185; PLASMID-PX01;
RA MEDLINE=20359347; PubMed=10899854;
RA Cohen S., Mendelson I., Altboum Z., Kobler D., Elhanany E., Bino T.,
RA Leitner M., Inbar I., Rosenberg H., Gozes Y., Barak R., Fisher M.,
RA Kronman C., Velan B., Shafterman A.;
RA "Attenuated nontoxinogenic and nonencapsulated recombinant Bacillus
RA anthracis spore vaccines protect against anthrax.";
RA Infect. Immun. 68:4549-4558(2000).
```

```
RL [2]
RL SEQUENCE FROM N.A.
RL PLASMID-PX01;
RL MEDLINE=99214082; PubMed=10197996;
RL Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RL "Genetic diversity in the protective antigen gene of Bacillus
RL anthracis.";
RL J. Bacteriol. 181:2358-2362(1999).
```

```
RN [3]
RN SEQUENCE FROM N.A.
RN PLASMID-PX01;
RN Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RN Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN EMBL: AF268967; AAF6457.1; -
RN EMBL: AF306782; AAG24450.1; -
RN InterPro: IPR003896; Binary_toxB.  
DR PRINTS: PR01391; BINARYTOXINB.  
KW Plasmid.  
SO SEQUENCE 764 AA; 85838 MW; B0DAFC1DCF83DAF4 CRC64;
```

Query Match 96.9%; Score 3793; DB 2; Length 764;
Best Local Similarity 97.4%; Pred. No. 5.6e-208;
Matches 744; Conservative 3; Mismatches 17; Indels 0; Gaps 0;

```
QY 1 MKKRVLLPMAALSTIIYSSGNGNEVIAEYKQENRLNLSSESSSGGLGYFSDLNQA 60
DB 1 MKKRVLLPMAALSTIIYSSGNGNEVIAEYKQENRLNLSSESSSGGLGYFSDLNQA 60
QY 61 PMVYTSSTTGDISPSSSELENIPSENQYFQSAIMSGFTIKVKSDEYTPATSDNHVTMW 120
DB 61 PMVYTSSTTGDISPSSSELENIPSENQYFQSAIMSGFTIKVKSDEYTPATSDNHVTMW 120
QY 121 DDQEVINKASNSNKTIRLEKGRLYQIKIOYORENPTKEGLDFKLWYTDSONKKEVISSDNL 180
DB 121 DDQEVINKASNSNKTIRLEKGRLYQIKIOYORENPTKEGLDFKLWYTDSONKKEVISSDNL 180
QY 181 QLPETKOKSSNRKRRTSAGPTVPDRNDGIPDSLEVEGYVDVKNKRTFLSPWISNH 240
DB 181 QLPETKOKSSNRKRRTSAGPTVPDRNDGIPDSLEVEGYVDVKNKRTFLSPWISNH 240
QY 241 EKKGLTKKSSPEKMTASDPYSDPEKTYGRIDKNVSEARHPPLVAAPYIVHDMENITL 300
DB 241 EKKGLTKKSSPEKMTASDPYSDPEKTYGRIDKNVSEARHPPLVAAPYIVHDMENITL 300
QY 301 SKNEDOSTQNTDSETRTISKNTSTSRHTSEVHGNAEYHANTSTSRHTSEVHGNAEYHA 360
DB 301 SKNEDOSTQNTDSETRTISKNTSTSRHTSEVHGNAEYHANTSTSRHTSEVHGNAEYHA 360
QY 361 VAIDHSLSLAGERTAEETMGNTADTARLANIRVNTGTAPYVAVLPTTSLVLCGNQTL 420
DB 361 VAIDHSLSLAGERTAEETMGNTADTARLANIRVNTGTAPYVAVLPTTSLVLCGNQTL 420
QY 421 ATIRAKENQSOIILAPNNYPSKNLAPIALNADDFSSPTITMANNQFLEKTKQLRLD 480
DB 421 ATIRAKENQSOIILAPNNYPSKNLAPIALNADDFSSPTITMANNQFLEKTKQLRLD 480
QY 481 TDQVYGNATYTNFENGRRVVDGTSNMSEVLPOIOETTAIIIFNGKDLNVERRIAANPS 540
```

```
Db 481 TDQYGNATATYFNENGRVVDGSMSEVLPOIOETTARIIFNGKDLNVERRIAANPS 540
QY 541 DPLETTPDMTLKALKIAFGNEPNNGNLQYQGDITEFPDNFPDQOQSQNTKNOLAELNA 600
Db 541 DPLETTPDMTLKALKIAFGNEPNNGNLQYQGDITEFPDNFPDQOQSQNTKNOLAELNA 600
QY 601 TNYITVLDKIKLNKMNILIRDKRFHYDRNNIAVGADESIVKKAHREYINSTEGLLNI 660
Db 601 TNYITVLDKIKLNKMNILIRDKRFHYDRNNIAVGADESIVKKAHREYINSTEGLLNI 660
QY 661 DKDIRKILSGYIVEIEDTEGLKEVINDRYDMLNITSLRQDGKTFIDFKKYNDKLPYISN 720
Db 661 DKDIRKILSGYIVEIEDTEGLKEVINDRYDMLNITSLRQDGKTFIDFKKYNDKLPYISN 720
QY 721 PNKVVNYAVTKENTIIINPSENGDTSTNGIKKILIFSKKGYEIG 764
Db 721 PNKVVNYAVTKENTIIINPSENGDTSTNGIKKILIFSKKGYEIG 764

RESULT 3
Q9F5R7 PRELIMINARY; PRT; 764 AA.
ID Q9F5R7;
AC Q9F5R7;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PROTECTIVE ANTIGEN.
GN PAG.
OS Bacillus anthracis.
OC plasmid pX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BA1024;
RX MEDLINE=99214082; PubMed=10197996;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RT "Genetic diversity in the protective antigen gene of Bacillus
RT anthracis."
RL J. Bacteriol. 181:2358-2362(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BA1024;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF306783; AAG24451.1; -.
DR InterPro: IPR003896; Binary_toxb.
DR PRINTS: PR01391; BINARYTOXINB.
KM plasmid.
SQ SEQUENCE 764 AA; 85828 MW; A1845CE1FEDCD93A CRC64;
```

Query Match 96.7%; Score 3785; DB 2; Length 764;
Best Local Similarity 97.3%; Pred. No. 1,6e-207;
Matches 743; Conservative 3; Mismatches 18; Indels 0; Gaps 0;

```
QY 241 EKKGLTKKSSPEKMWSTASDPYSDPEKYVGRIDKNVSEARHPLVAAYPIHVDMENITL 300
Db 241 EKKGLTKKSSPEKMWSTASDPYSDPEKYVGRIDKNVSEARHPLVAAYPIHVDMENITL 300
QY 301 SKNEDQSTONTSETRTISKNTSTSRHTSEVHGNAEVHANTSTSRHTSEVHGNAEVHA 360
Db 301 SKNEDQSTONTSETRTISKNTSTSRHTSEVHGNAEVHANTSTSRHTSEVHGNAEVHA 360
QY 361 VAIDHSLSLAGEERTWAEIEMGLNTADTARLANINRYVNTGAPITVNLPTTSVLGKNOTL 420
Db 361 VAIDHSLSLAGEERTWAEIEMGLNTADTARLANINRYVNTGAPITVNLPTTSVLGKNOTL 420
QY 421 ATRAKENQSLQIAPNNYPSKNLAPALNAODFSEPTTMMVNOFLELEKTKOLRLD 480
Db 421 ATRAKENQSLQIAPNNYPSKNLAPALNAODFSEPTTMMVNOFLELEKTKOLRLD 480
QY 481 TDQYGNATATYFNENGRVVDGSMSEVLPOIOETTARIIFNGKDLNVERRIAANPS 540
Db 481 TDQYGNATATYFNENGRVVDGSMSEVLPOIOETTARIIFNGKDLNVERRIAANPS 540
QY 541 DPLETTPDMTLKALKIAFGNEPNNGNLQYQGDITEFPDNFPDQOQSQNTKNOLAELNA 600
Db 541 DPLETTPDMTLKALKIAFGNEPNNGNLQYQGDITEFPDNFPDQOQSQNTKNOLAELNA 600
QY 601 TNYITVLDKIKLNKMNILIRDKRFHYDRNNIAVGADESIVKKAHREYINSTEGLLNI 660
Db 601 TNYITVLDKIKLNKMNILIRDKRFHYDRNNIAVGADESIVKKAHREYINSTEGLLNI 660
QY 661 DKDIRKILSGYIVEIEDTEGLKEVINDRYDMLNITSLRQDGKTFIDFKKYNDKLPYISN 720
Db 661 DKDIRKILSGYIVEIEDTEGLKEVINDRYDMLNITSLRQDGKTFIDFKKYNDKLPYISN 720
QY 721 PNKVVNYAVTKENTIIINPSENGDTSTNGIKKILIFSKKGYEIG 764
Db 721 PNKVVNYAVTKENTIIINPSENGDTSTNGIKKILIFSKKGYEIG 764

RESULT 4
O32739 PRELIMINARY; PRT; 876 AA.
ID O32739;
AC O32739;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ADP-RIBOSYLTRANSFERASE.
GN CDTB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CD196;
RX MEDLINE=97230316; PubMed=9119480;
RA Perelle S., Gilbert M., Bourlioux P., Corthier G., Popoff M.R.;
RT "Production of a complete binary toxin (actin-specific ADP-
RT ribosyltransferase) by Clostridium difficile CD196."
RL Infect. Immun. 65:1402-1407(1997).
DR EMBL: U76081; AAB67305.1; -.
DR HSBP: P13423; IACC.
DR InterPro: IPR003896; Binary_toxb.
DR PRINTS: PR01391; BINARYTOXINB.
KM transferase.
SQ SEQUENCE 876 AA; 98796 MW; 25E06E2D45CE2B3B CRC64;
```

Query Match 23.0%; Score 900; DB 2; Length 876;
Best Local Similarity 32.1%; Pred. No. 3.1e-43;
Matches 268; Conservative 145; Mismatches 285; Indels 136; Gaps 36;

[illegible]

RT CCUG 20309, " to the EMBL/Genbank/DBJ databases
 RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases
 DR EMBL: AF271719, AAF81761.1; -
 DR InterPro: IPR003896; Binary.t0x.B.
 DR PRINTS: PR01391; B1NRYOXI0XB.
 SO SEQUENCE 876 AA: 98792 MW: 366D62F352E745A5 CMC6;

Query Match	22.9%;	Score 898;	DB 2;	Length 876;
Best Local Similarity	32.0%;	Pred. NO. 4e-43;		
Matches 267;	Conservative 146;	Mismatches 285;	Indels 136;	Gaps 36;

[illegible]

RESULT	6	
006498		
ID	006498	PRELIMINARY;
AC	006498;	PRT; 879 AA

DT 01-JUN-1997 (Tremblrel. 04, Created)
 DT 01-JUN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE SP COMPONENT.
 GN SRS.
 OS Clostridium spiroforme.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes.
 NCBI_TaxID=29348;
 RN [1]
 RP SQUENCE FROM N.A.
 RC STRAIN-CS246;
 RA Gilbert M., Perelle S., Daube G., Popoff M.R.,
 RL Syst. Appl. Microbiol. 20:337-347(1997).
 DR EMBL: X97969; CA66612.1; -
 DR HSSP: P13423; IACC.
 DR InterPro: IPR003896; Binary.toxb.
 DR PRINTS: PRO1391; BINARYTOXINB.
 SQ SEQUENCE 879 AA; 98738 MW; 40685ACB8E05BA01 CRC64;

Query Match 22.6%; Score 886; DB 2: Length 879;
 Best Local Similarity 28.6%; Pred. No. 2e-42;

Matches 270; Conservative 151; Mismatches 270; Indels 252; Gaps 37;

QY 1 MKKKVLPIMALSTILVSSTGNLEVIQAEV-----KOENRLNSESSESSOGILGYFSD 55
 DB 1 MKKKKILGLTC--TVLVGOMMTYPPYAKITITONVNDQEVETNEKTVSSGILMGYTFAD 58
 QY 56 LNFQAPNVVTSSTTGDILSPSELENIPISENQ-YFQSAIWSGFIKVKSDSEYTFATSADN 114
 DB 59 EHEKDELIMAPVKNGELFKFNKVEKLETKETNIKSIHWGRIIPSKDGEYTLSTDKDN 118
 QY 115 HYTMVDDOVINKASNSKIRLEKGLYQIKYOORENTEKGLD-----KLTWTD 167
 DB 119 -VLMQINAE--GEIANTLVNMKGGQYSIRLEIQ-----DKDIGVDLSPEKLV-E 168
 QY 168 SQKKKEVISSDNLQPELKOKSSN-----SRKKRSTSA-GPTVPRDNDG 211
 DB 169 LNDGKTLIPKKNLFLRYSKSIDENDPRPKDNFDLKLKRSARLASKGWDDELDTNDN 228
 QY 212 IPOSLEVEGYTVDKNKRFLSPWISNIHEKGLITKXSSPEKWSASDPYSDEKVTGR 271
 DB 229 IIPAYEKNGTYI---KDSIAVW-EDSFAQGGYKKYLSYLSBSNAGDEYTDQKASGS 283
 QY 272 IDKNVSEARHPLVAAPYIVHVMENILSKNEOSQNTDSEKRTSKMTSTRT--HT 329
 DB 284 FDAIKAEADPLVAAPYVGVGHEKLIISTNEAST---DQKTVSRMTTNSKTANT 339
 QY 330 SEVHNAEVAH---ANTSTSTSEVHNAEVAHVAIDHSLSLAGERVAETWGLTAD 385
 DB 340 AGVAINAVNGFTGSTITYSHTE-----NSTAVONS--NGE-SMNTSLSTIKGE 388
 QY 386 TABLANIRYVNTGTAIVVLEPTSLVLGKNOTLATIKAKENQLSQILAPNNYPSKNL 445
 DB 389 SAYINANVRYNGTAPMYKYVPTTNLVL-DGDTLTITIKADQNGNLSPNETYPKGL 447
 QY 446 APALNAADDESSPTITMNTNOLELEKTKQLRDTDOVYGNATYFENGRAVVD-GS 504
 DB 448 SPLALTMDFSSRLPIINDOLKLDAGKQIKLETTQVSGN--YQIKNSOGQIITEGN 504
 QY 505 MNEEVLPOIETARTIIFN-GKDLNVERLIAVNPSPDETTRKPDMLTEAKIARGFN 563
 DB 505 SMSDYISOIDSLSASITILDGSD--VEFRKYVAKDSSNPEDKT-PVLTIGDALEKAGAT 561
 QY 564 EPNGLNOYOGKDIITE--FDENFDQOSQNTKQNLQELANATNYTVLADKILNKMNLIR 621
 DB 562 KNGEILYFNMGPIDESCVELIFDGNANTLKERLMAINDKRIYV---QLERKMLIK 617
 QY 622 D----- 622
 DB 618 TSTYFNFDGYNPSSMSVNSNODGLQMANAKLSGETKIYIPMSKLNPKRYVESGY 677
 QY 623 -----KRFHY-----DRNN----- 631

DB 678 LKNSTSNPIVNTIKAKEQKTYNLVSENDYKFFSEFEPTIGROASNIETLTSSGTIFLD 737
 QY 632 -----IAGADESVYKAEHREY-----INSTEGILLN-----I 660
 DB 738 NLSITELNSPELLEKPIDKIPSDQDEII-DAHKRYVADLSFNQSTANVYLDGLYFEPLOT 796
 QY 661 DKDIRKILSGYVIEI-DTEGLKEVINDRYDMLNSSLROGKT-FIDFKKNKPLXI 718
 DB 797 NKEVLDTQKIKVATLEYSGFKDGTGKDKELRNTYDGSNOPTNYNFRSY-----FT 850
 QY 719 SNPNY-----KVNVAVTKENTINPSENGDSTNGIKKILFS 757
 DB 851 SGENVMPYKKRIYAITPEN-----KELVLVS 877

RESULT 7
 ID Q46221 PRELIMINARY; PRT; 875 AA.

DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE IOTA TOXIN COMPONENT IB PRECURSOR.
 OS Clostridium perfringens.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 NCBI_TaxID=1502;
 RN [1]
 RP SQUENCE FROM N.A.
 RC STRAIN-NCIB 10748;
 RX MEDLINE=94041637; Pubmed=8225592.
 RA Perelle S., Gilbert M., Boquet P., Popoff M.R.;
 RT "Characterization of Clostridium perfringens iota-toxin genes and
 RL Infect. Immun. 61:5147-5156(1993).
 RN [2]
 RP SQUENCE FROM N.A.
 RC STRAIN-NCIB 10748;
 RA Popoff M.R.;
 RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SQUENCE FROM N.A.
 RC STRAIN-NCIB 10748;
 RA Popoff M.R.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X73562; CA51960.1; -
 DR HSSP: P13423; IACC.
 DR InterPro: IPR003896; Binary.toxb.
 DR PRINTS: PRO1391; BINARYTOXINB.
 KW Signal.
 FT SIGNAL 34
 FT CHAIN 212
 FT SEQUENCE 875 AA; 98468 MW; C9AE092CD3818921 CRC64;

Query Match 22.6%; Score 883.5; DB 2: Length 875;
 Best Local Similarity 31.5%; Pred. No. 2.7e-42;
 Matches 262; Conservative 135; Mismatches 279; Indels 155; Gaps 35;

QY 15 TLVSSSTGNLEV-----IQAEVKQENRLNSESSESSOGILGYTSIDLNFQAPNVVTSST 69
 DB 15 TMLTSQTLSTYVNAQTTOQNTDNTNOKREBITMENTLSSNGLMGYFADHFDLELMADIKN 74
 QY 70 GDLISPESELENIPSE-NOYFQSAIWSGFIKVKSDSEYTFATSADNHYTMVDDQEVYINK 128
 DB 75 GDLKFEKKYDKLITEDNSSIKSTIRWTGRIIPSDGEYIISTDR-NDVLMQINKAGDIK 133
 QY 129 ASNSNKIRLEKGLYQIKYOOREN-----PTEKGLDFKLYWTDSONKKEVISSDN 179
 DB 134 ---TLKVMKKGQAYNIRIEIODKSLIDMLSYV-----KLYW-ELNGNKTIVPEEN 182
 QY 180 LQPELKOKSSNSRKRSTSGPIVP-----DADNDGIPDSLEVEG 220

```

Db 183 LFFRDYSKIDEND-----PFIPINNFEDVREFSAMEDDEDLDTNDNDIPDAYEKNG 233
Oy 221 YTVADKNRKTRFLSPMISNIHEKKGLTKYKSSPEKSTASDPYSDEKATGRIDKIVSPEA 280
Db 224 YTI-----KDSIAVKNKNDSPAE-QGKTKYVSYLESTADDPYTDYOKASGSIDKAIKLEA 288
Oy 281 RHPLEAAVPIVHVDMENTILSKNEOSTONTDSETRTISKNTSRTHSEVHGAEEVA 340
Db 289 RDPLEAAVPIVHVDMENTILSKNEOSTONTDSETRTISKNTSRTHSEVHGAEEVA 342
Oy 341 -----NTSIRTHSEVHGAEEVAHVAIDHSLAGERTMAETMGLINTADRLAN 392
Db 343 GYONGFTGNITTSYHTTD-----NSTAVODS---NGE-SWNTGLSINKESAYIAN 391
Oy 393 IRYVNTGAPITVNLPTTSYLVEKNOTLATIKAKENOSQILAPNNYPSKRLAPITANA 452
Db 392 VRYVNTGAPITVNLPTTSYLVEKNOTLATIKAKENOSQILAPNNYPSKRLAPITANA 450
Oy 453 ODDFSTPTMNYNOFLELEKTRKQLRLDTPDYGNIAIYFENGVRVDTGSMSEVLPO 512
Db 451 MDQFARLPIPIVYDQKLDKLSGKQIKLETTQVSGVNGTKN-SQGOI-ITEGSMSEYISO 508
Oy 513 IOETARIIFNKDNLIVERIAAVNSDPLETKPDMTKALKATIAFGFNPENLOY- 571
Db 509 IDSVASIILD-TGSQTEERRVAAKEQGNPEDKT-PEITIGAIKAKASATK-NEGLLYE 565
Oy 572 QGKDIETE--FDFNPDQOTSQNIKNQLAELMNTIYVLDKIKLNKMMILDRKRF--HY 627
Db 566 NGIPIDESCVELIFPDNTNSEIIEKOLKYLDKKIYN-----KLEGMNLLIVPSYFTNF 621
Oy 628 DR-NNIAVGADESVYKAEHREVIANSSTGL-----LNLIDDKIRKILSGY----- 671
Db 622 DEYNMF--ASWSNIDTKQMDGLQSVANKLSGETKIIIPMSKLPYKRVFSGSKDPT 679
Oy 672 -----IVEIEDTEGLKEVINDRYDMLNIS-----SLROGKTIDSKKYND 712
Db 660 SNSITVNKSKSQKTDYLVPEKDYTKFSEFETTKDSSDIEITLTSSGVIFLDMLSITE 739
Oy 713 --KLPLYISNPNYKV-----NYAVTKENTIIINSENGPTSTNGI 750
Db 740 LNSTEILKEPEIKVPDSOEILDAHKNKYADIKLDT-----NTGNTYIDGI 785

```

RESULT 8

086171 PRELIMINARY: PRT: 721 AA.

086171:

AC 086171:

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)

DE C2 TOXIN (COMPONENT-11).

OS Clostridium botulinum.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;

CC Clostridium;

OX NCBI_TaxID=1491;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-TYPE C (C)-203028;

RX MEDLINE=96323874; PubMed=9659689;

RA Kimura K., Kubota T., Ohishi I., Isogai H., Isogai E., Fujii N.;

RT "The gene for component-II of botulinum C2 toxin."

RL Vet. Microbiol. 62:27-34(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-TYPE C (C)-203028;

RX MEDLINE=96184657; PubMed=8645309;

RA Fujii N., Kubota T., Shitakawa S., Kimura K., Ohishi I., Morishiki K.,

RT "Characterization of component-I gene of botulinum C2 toxin and PCR

RL detection of its gene in clostridial species."

RM Biochem. Biophys. Res. Commun. 220:353-359(1996).

EMBL: D88982; BAA32537.1; -.

```

DR HSSP; P13423; IACC.
DR InterPro: IPR003896; Binary_toxb.
DR PRINTS: PR01391; BINARYTOXB.
SO SEQUENCE 721 AA; 80515 MM; 44C8153AC749D5F2 CRC64.

Query Match 20.4%; Score 798; DB 2; Length 721;
Best Local Similarity 34.5%; Pred. No. 1.5e-37;
Matches 221; Conservative 117; Mismatches 221; Indels 82; Gaps 25;

Oy 32 KOENRLNESES--SSQGLGYFESDLNFQAPMVYSSTTGDLSPSSLENIPESENOYF 89
Db 5 KFEVSKNSKNKYFTINGIMGYFEN-DEFNLNIIISPTLDGTFPSKEDINSILG-NKII 62
Oy 90 OSATWSGFYKVKKSDYEYFATSADN-HYTMWDDQDEVINKASN-SNKIRLEGRGYOKI 147
Db 63 KSARHIGLIKPSITEGYELISTNSPCRELB--NEIRNLSINTNYNLIOGNVYDIRI 119
Oy 148 -QYRENPTEKGLD-FKLWTDSONKKEVISSDNLQLPKOKSSSKRSTSGPTV- 204
Db 120 EQLMSENOLKNYEGIKLYWETSIDIKETIPEVL-----LKPNTSEKSKFIPNNTLF 175
Oy 205 -----PDRNDGTPDLSLEVGYVDYKNNKTFELSPWISNIHEKKGLTKYKSSPEK 254
Db 176 SNAKLKANNRDTDRDGIPEDEWEINGYVMNOKAVAMDKEFAN-----GIKTKYSNPFK 230
Oy 255 WSTASDPYSDEKRYGRIDKNVSPARHPPLVAAPIVHVDMENTILSKNEOSTONTDSE 314
Db 221 PCTANDPTDFEFGVGOIDPSVSMVARDPMISAYPIVGYOMRLVYKSE-----TITGDS 286
Oy 315 TRTISKNTSTSKTHF-----SEVHGAEEV-----HANSTSRTHSEVHGAEEV 358
Db 287 TKMSKSTSHSSTNTNITGAEVSGSLQLAGIFPVFSMASASANYSHTQNTSTVD----- 341
Oy 359 HAVAIDHSLAGERTMAETMGLINTADRLANLRYVNTGAPITVNLPTTSYLKNG 418
Db 342 -----DTTGE--SFQGLSINTGESAYINPNRIYNTGAPVYNTPTTYIDV-Q 390
Oy 419 TLATIKARENOLSQLAPNNYPSKNLAPIALNADDDSTPTMNYNOFLELEKTRKOLR 478
Db 391 SVATIKQGESLGDVLNCGTPIIIGPPMALNTMDQSSRLRIPNTYNOLKSIDNGCYVM 450
Oy 479 LDTQVYGNIAIYFENGVRVDTGSMSEVLPOIETTARIIT--FNKDLNVERIAA 536
Db 451 ISTSOTGTFAYYN-SNGVLVTD-GNNMGPLYGTIKSTASLTLSFSGCQTQOVA--VVA 505
Oy 537 VNPSPLETTKPDMLKEALKIAPFNENGLQVGGDIT--FEFDFNPDQOTSQNIKN 593
Db 506 PNFSDPEDKT-PKLTLEQALVKAFALEKKNKGFYHGLEISKNEKIOYFLOSNTNDFEN 564
Oy 594 QIAELNATNIYVLDKIKLNKMMNLLIRDRKREHYDRNNIAV 634
Db 565 QKNTADMDIMHCI--IKRN--MNLVAVYTFEKNISSINI 601

```

RESULT 9

09X377 PRELIMINARY: PRT: 204 AA.

09X377:

AC 09X377:

DT 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)

DE PX01-111.

OS Bacillus anthracis.

OC Plasmid virulence plasmid PX01.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=1392;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-STERNE.

RA Okinaka R.T., Cloud G., Hanton O., Hofmaster A., Hill K.K., Keim P.,

RM Koehler T., Lamke G., Kumano S., Mahillon J., Mantler D., Martinez Y.,

RA Rieke D.O., Svensson R., Jackson P.J.;
RT "The sequence and organization of pXol, the large Bacillus anthracis
RL plasmid harboring the Anthrax toxin genes.";
DR EMBL: AF065404; 0-0-0(1999).
DR HSSP: P13423; IACC.
KW Plasmid.
SQ SEQUENCE 204 AA; 23029 MW; E1657B23AE4273FD CRC64;

Query Match 6.0%; Score 235.5; DB 2; Length 204;
Best Local Similarity 34.6%; Pred. No. 2,7e-06;
Matches 56; Conservative 36; Mismatches 47; Indels 23; Gaps 6;

QY 616 MNLIRDRKRRHYDRNNANVGADESVKAEHREYNSTEGLLMDIDIRKILSGYIEI 675
DB 1 MNLIVRDP-YHYDNGNIVGVDSDYLKNAKQILMSSDGLMLDDVDNALSGYMIQI 59
QY 676 EDTE-----GLKEVINDRYDMLNSSLRODGKFTIDFKRYNDKLPYISNP 722
DB 60 KKSNNHLTNSPVITTLGKRSQSGVELRYVLS-----DGICFLDPKPFDEWMSLV-DPG 112
QY 723 YKVVYAVTKEN-TIIPSENGDTSTNGIKKILFSSKGYEI 763
DB 113 DQVYVAVTKEDFNVAVTRDENGNI-NKLNKTLVLSGKIREI 153

RESULT 10
Q9P008 PRELIMINARY; PRT; 4688 AA.
AC Q9P008;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN U0482.
GN U0482.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacilli/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID-134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROVAR 3;
RX MEDLINE-20500219; PubMed-11048724;
RA Glass J.I., Letkovitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum.";
RL Nature 407:757-762(2000).
DR EMBL: AE002145; AAF30894.1; -
DR InterPro: IPR001152; Thymosin_b4.
DR InterPro: IPR000626; Ubiquitin.
DR Pfam: PF00240; ubiquitin.1.
DR SMART: SM00152; TRY.1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 4688 AA; 534880 MW; B53ABFAFEEL1997E CRC64;

Query Match 5.3%; Score 205.5; DB 2; Length 4688;
Best Local Similarity 20.8%; Pred. No. 0.011;
Matches 176; Conservative 138; Mismatches 337; Indels 197; Gaps 42;

QY 18 VSTGNTLEVI-QAEVKQENRLNSESOGGLGYFSDLNFQAPMVVYTSSTGDLSTPS 76
DB 3676 INTGKQEVLEFVSKGLSNOLYK-----LVVVYILD-NIHONIDETKRIFFDHNV-S 3725
QY 77 SELEINISENQYFOSALISGFIKVKSDSEYTFATFSADNHNVMVDDEVINKASNSKIR 136
DB 3726 KEIFINGVMIMSKHGKMSKPTDTTANFERKIEFYQ-----DDNDVLNNIDATYAEK 3776
QY 137 LEKRLQYIKIYOYRENPTEKGLKLYWTDSONKKEVSSDNLQLELKQSSN---S 192
DB 3777 DEHNINIKQIKYIRIKEN-----NDMLIKQI---DNLN-PETKKLENIELSK 3820

QY 193 RKRSTSGAPVPPDR-----NDGIPDSLVEGYVDVKNKRTFLS 233
DB 3821 PLKTHNLNLSVINDKENISLITTEGNPLKAYIQUNDINTOQTINTVLSGVSK-YNG 3879
QY 234 PWISNHEKKGLTRYKSSPEKWSASDPYSDFEKVTGRIDKNVSPKARHPLVAAYPIVH 293
DB 3880 RQIKVYKADNNNVNVESS---LITLQKGKNDYQLLSLNLSN-----REYFEKEIETHI 3931
QY 294 -----DME-----NIIISKNEQSTQNTDSE-----TETISKNSTSTHSEVGN 335
DB 3932 SNTNNEFDELEKNGVSNFTIQTKNITVQUNDSSATIGTGVGNFENK1-KSEDKILENN 3990
QY 336 AEV---HANTSTSRTHSEVGNAEVNAHVAIDHSLSLAGEPTMETGLNT---ADARL 389
DB 3991 QQVVAWEPKKEITIDTNTWLOAYTRPLKDVTSDFK-----EGTMAHDSNSNFEKEETTK 4045
QY 390 NANIRVYVTGAPLYNVLPPTSLVGLKNQTLA-----TIKAKENQSLIAPNNVYPSK 443
DB 4046 LKIQFVKKPTKAKNNINNSNNVILDTNTINSNVEETTKVGDHKLINITSNNVNTNS 4105
QY 444 NLADIALN-AODFSSPEITVNY---NQFLEKTKQLRDTDOYGNATYFENGK-- 497
DB 4106 QTIHFETLSGVKKSWMGKKIKLSYKSDNTSESITHNEVLEESNKTQYNILNKLKRNFTYT 4165
QY 498 ---VRVDGSMWSEVLPOIQTETARIIFNGKDLNVERIAVAPSDPLETT-----K 547
DB 4166 LIDVKLIDNNVSDPPEKGNLTNSFTTTSAINVLNIEISNRASNTAKSTIIKININD 4225
QY 548 PDMTLKEA-----LKIAGFNEPKNLOYOGKDTTEPFEDQOTSONIK-NOLAE 597
DB 4226 PDNVLRDDQATIIYGNKKQAMGFIYSGNIKILATIVDLNPN-DKNYINISFNKPS 4284
QY 598 LNAITNITVLDKIKLNAMKMLINDKRFHYD-----RNNIAYGA---DESVKAEHRE 647
DB 4285 IAAEN-----IGIDKSNNTI-----YNDISIPKLEINNDIIVGPIKEIYVKNAMOK 4332
QY 648 VINSSTEGLLINDKDIRKILSGYIELEDEGLKEVIN-D-RVDMILSSL-RQDGTF 704
DB 4333 --NNIDVDLGIQINKIAHNLR-FLAKFKSTN-----NDIIEFNVLIGSSLVNNDGTF 4383
QY 705 IDF-----KKY-----NDKLPYISNPYKVVAVTAKENTIIIN 738
DB 4384 IRLFLNNKAKKLISLVVYVYLVNNSNTIYESNKL-P-KLNNINQIK---INSHITII 4438
QY 739 PSENGDTS 746
DB 4439 -SKNGEMS 4445

RESULT 11
Q9G276
ID Q9G276 PRELIMINARY; PRT; 1387 AA.
AC Q9G276;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE REFICULOCYTE-BINDING PROTEIN 2 HOMOLOG A (FRAGMENT).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID-5633;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20402589; PubMed-10920203;
RA Rayner J.C., Galinski M.R., Ingravallo P., Barnwell J.W.;
RT "Two Plasmodium falciparum genes express merozoite proteins that are
RT related to Plasmodium vivax and Plasmodium yoelii adhesive proteins
RT involved in host cell selection and invasion.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:9648-9653(2000).
DR EMBL: AF196347; AAF98066.1; -
FT NON_TER 1 1
FT NON_TER 1387 1387
SQ SEQUENCE 1387 AA; 164871 MW; 5660154744F0C285 CRC64;

Db 1708 MDGPISNVNPEELHPVAEGSKLEAKERSMDADKQ---TTEEDIT-----VVEDPNC 1758
QY 675 IEDTBELKEVINDRYDMLNSSLROGKPTIDDKKNDKPLIISNPYVNVYAATKEN 734
Db 1759 IGHQNLKEVHEQASBELMNTYNSL--DGRINVEVKERLDENPGSIPDRITTEHIEIDREK 1816
QY 735 TIINPSE---NGD 744
Db 1817 EIEHPNELDAHNGE 1830

RESULT 13
Q9BK45 PRELIMINARY; PRT; 3254 AA.
AC Q9BK45:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE RETICULOCYTE BINDING PROTEIN 2 HOMOLOG B.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21101060; PubMed=11160005;
RA Triglia T., Thompson J., Caruana S.R., Delorenzi M., Speed T.,
Comman A.F.:
RT "Identification of Proteins from Plasmodium falciparum That Are
RT Homologous to Reticulocyte Binding Proteins in Plasmodium vivax";
RL Infect. Immun. 69:1084-1092(2001).
DR EMBL: AF312917; AAK19245.1; -
SQ SEQUENCE 3254 AA; 382876 MW; 6F9CAFA5AA6167BA CRC64;

Query Match 5.0%; Score 195.5; DB 5; Length 3254;
Best Local Similarity 19.2%; Pred. No. 0.024;
Matches 172; Conservative 140; Mismatches 294; Indels 291; Gaps 41;

QY 5 KVLIPALSTIIVSSGTGNEVIAEQAEVKQENRLNSESSESSGCLGYFSDLNFOAPMV 64
Db 1920 KHVYIKLANFSGIIVMSDTNTEITPENPLENDLLN-----LQLYFERKHEI 1965
QY 65 TSSGTGDLSPSSELENI--PSENOYFQSAIWSGFVKKSDGYTFATSDAHVTMVAVD 122
Db 1966 TSTLENDSDL--ELDHLSGNSDESIDNLKVYNDIETL-----HYST-----QILKYLDN 2013
QY 123 -----OEVIKASN-SNKIRLEKGRLYOI-KIQ 148
Db 2014 IOKLKGDCNDLVKDEKRELSTALYDLKIQITSVINRENDISNNIDIVSNKLINEIDAIO 2073
QY 149 YORE-----NPTKGDGDFKLYMTDSQNKKE 173
Db 2074 YNEEKKEIFDNEVEYKTLDDTKNAVYKAEILKANVDINKTKTEDDIDYFNDDLEKSL 2133
QY 174 VISSDNLQPELKQSSNRKRSSTAGPTVPDRNDG---IP--DSLVEGYTVD----- 224
Db 2134 TSSNMEIKTIYQNSYNS--FSDINKNINDIKEMKTLIMDLBELNEGHNIDISLY 2189
QY 225 ---VKNKRTFLSPWISNIHEKKGLT---KYKSSPEKMWSTAS---DPYSDFEKGVRID 273
Db 2190 NFILRNIOIKIGNDIKINIKREQENDTNICEFYIGNNNYFIKSDISIFNKYDDHKKVNYIS 2249
QY 274 KAVSPARPLVAAPYIVVD--MENI---ILSKND-----OSTO-----NTDSE 314
Db 2250 NNIDVYNNKNSLSLSEHIVNATNIENIMTSIYINEDTEMSLLETOQDLLELYENPKKE 2309
QY 315 TRTISKNTSTSR--TRTSEVHGAENVAHANTSTRTSTSEVHG-----NAEV 358
Db 2310 KNINNNYIIVHENKLEIENSLLETYNSTINKNKINETONIDILKNERNITKTKNDKY 2369
QY 359 -HAVALDHSLSLAGERTMAETMG--LNTADTARL---NANIRYVNTGTAPYVNLPTT 410
Db 2370 KELVHVDSTLTLESIGTFNNILYGDLMNSIQDYYVKYEDINNVELKVKYKLYIENTNTLIGRI 2429

QY 411 SLVIGKNOGLATITKAKENOLS---QILAPNNYPSKRLAPIALNAODESPSTPTMANYQ 467
Db 2430 NTEFI---KELDKYQDENNGGIDKYEIENKKNNSYIIR-LKEKANLKNFSS-----K 2476
QY 468 FLELEKTKQLRLDDQYVGNIAATYFENGVRVDGTSNMWSEVLPIQIETARIIFNGKDL 527
Db 2477 LLONIKRNTEFL-----YNINN--IKDD-----IMNTGKSV 2505
QY 528 NLVERRIAANVPSDPLEETTRKPDMLTKALKIAGFENPENGNOLOYGKNDITFDFPNFDOOT 587
Db 2506 NNIKQKFFSS-----NLPLKEKL--FQMEEMLANI----- 2532
QY 588 SONIKNOGLAELNATNTYTVL-----DKIKLAKAMIIILROKRF-HYDRNNIAVGADES 639
Db 2533 -NNIMNETKRISNTDATTNTTLQDIENNNKKNENNNMIETIDKIDIKIKHNEIQAEIL 2591
QY 640 VKEAHREVINSSTEGILLNDKIDRKILSGYIYEIDTEGLKEVINDRYDMLNSSLRQ 699
Db 2592 IIDDARKKV-----KETTDINKAFNEITENYNN-----NNGVIR 2627
QY 700 DGKTFIDFKRY-NDKLPYISNPYKVYVAVTKENTINPSENGDSTNGIKKILI 755
Db 2628 SAKNIYDVAKATYLNNELDKFL-----LKNELLS-HNNNDIKDGDDEKLIL 2671

RESULT 14
Q9BK46 PRELIMINARY; PRT; 3130 AA.
AC Q9BK46:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE RETICULOCYTE BINDING PROTEIN 2 HOMOLOG A.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21101060; PubMed=11160005;
RA Triglia T., Thompson J., Caruana S.R., Delorenzi M., Speed T.,
Comman A.F.:
RT "Identification of Proteins from Plasmodium falciparum That Are
RT Homologous to Reticulocyte Binding Proteins in Plasmodium vivax";
RL Infect. Immun. 69:1084-1092(2001).
DR EMBL: AF312916; AAK19244.1; -
SQ SEQUENCE 3130 AA; 370415 MW; 13D973DB89D82026 CRC64;

Query Match 4.9%; Score 193.5; DB 5; Length 3130;
Best Local Similarity 19.7%; Pred. No. 0.03;
Matches 173; Conservative 142; Mismatches 310; Indels 253; Gaps 44;

QY 5 KVLIPALSTIIVSSGTGNEVIAEQAEVKQENRLNSESSESSGCLGYFSDLNFOAPMV 64
Db 1920 KHVYIKLANFSGIIVMSDTNTEITPENPLENDLLN-----LQLYFERKHEI 1965
QY 65 TSSGTGDLSPSSELENI--PSENOYFQSAIWSGFVKKSDGYTFATSDAHVTMVAVD 122
Db 1966 TSTLENDSDL--ELDHLSGNSDESIDNLKVYNDIETL-----HYST-----QILKYLDN 2013
QY 123 -----OEVIKASN-SNKIRLEKGRLYOI-KIQ 148
Db 2014 IOKLKGDCNDLVKDEKRELSTALYDLKIQITSVINRENDISNNIDIVSNKLINEIDAIO 2073
QY 149 YORE-----NPTKGDGDFKLYMTDSQNKKE 173
Db 2074 YNEEKKEIFDNEVEYKTLDDTKNAVYKAEILKANVDINKTKTEDDIDYFNDDLEKSL 2133
QY 174 VISSDNLQPELKQSSNRKRSSTAGPTVPDRNDG---IP--DSLVEGYTVD----- 224
Db 2134 TSSNMEIKTIYQNSYNS--FSDINKNINDIKEMKTLIMDLBELNEGHNIDISLY 2189

QY 225 ---VKNRFTLSPWISNHEKKGLT---KXSSPEKSTAS---DPYDPEKYTGRIID 273
DB 2190 NFIIRNKIOTKIGNDIKNREQENDPTICEFYIQNNYNNFKISDISEFNKYDHIKIDNYS 2249
QY 274 KNSPEARHPVLAAYPIVHV--MENI---ILSKNED---OSTONTSETRTISKNTSTS 325
DB 2250 NNIDVKNHNSLSEHYINANNIENITMTSTVEINEDPEMSLEETQOKLELEYENFKE 2309
QY 326 RTHSEVHNAEVAHANTSTSTHTSEVHNAEVAHAVALDHSLSLAGEPTMAETMGLMTAD 385
DB 2310 K---NIINNKKI-----VHFN---KLKEIENSL-----ETYN----- 2336
QY 386 TARTLANIRVYNTGAPRYNVLPTTSVLGKNQTLATIKAEQO-----LSQ 432
DB 2337 --SISTNNKIN-----ETONIDILKNE--FNNIKTRINDKVKELVHVSTLTLES 2383
QY 433 ILAANNYPSKNLAPIALNADDESSPTITMNYNOFLEKTKOLRDTDOVYGNATYN 492
DB 2384 IQTFNNILY-----GDMNSNIDYVKYEDINN---VELKKVKLYIENTITLGLRINTFI 2433
QY 493 FENGRVRVDGSSNNSEVLPOIOETTARII--FNGKDLNIVERRIAANPSDPLETTKPDMT 551
DB 2434 KELDKYQ--DENNGIDKYEIKENNSYIILKEKANNLKE-----NFSKLLQNIKRNET 2486
QY 552 LKEALKIAFGFNEPNGNLOYGKDTTEDEFNPDQTSO--IKNQLAE-----LNATNY 604
DB 2487 -----ELYNNINIKDIDMTGKSVN---NIKQFSSNLDPLKEKLFQMEMLNINIM 2536
QY 605 TVLDKIKLNAK--MNLIRD--KRFHYDRNNIYAGADESVREAHREYNSSTEGILLNID 661
DB 2537 NETRISTATAVYNTITLQDIENNNKKNNNNIETIDKLID--HIKIHNEIQAEILLID 2594
QY 662 KDIRKILSGYIEVEDTEGLEKEVIN--DRYDMLNSSLRODGKTFIDFKKY--NDKPLV 717
DB 2595 DAKRRVK-----EITDNIKAFNETEYENYNNEN--NGVKSANKIYDEATYLNNEIDKF 2646
QY 718 ISNPNKYVAVYTKENTIIINPSENGDSTNGIKKILI 755
DB 2647 L-----LKNLELLS--HNNNDIKLDGDEKLIL 2671
RESULT 15
09ZHL0
ID 09ZHL0 PRELIMINARY: PRT: 4919 AA.
AC 09ZHL0;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LARGE SUPERNATANT PROTEIN 2.
GN LSPA2.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_Taxid=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=35000.
RX MEDLINE=99030326; PubMed=9811662;
RA Ward C.K., Lumbley S.R., Latimer J.L., Cope L.D., Hansen E.J.;
"Haemophilus ducreyi secretes a filamentous hemagglutinin-like
RT protein".
RL J. Bacteriol. 180:6013-6022(1998).
DR EMBL: AF057696; AAC79761.1;
DR InterPro: IPR00130; Zn_MTPeptide.
DR SMART: SM00235; znmC; 1.
SQ SEQUENCE 4919 AA; 542602 MW; 5779201455CA69A0 CRC64;

Query Match 4.98; Score 193.5; DB 2; Length 4919;
Best Local Similarity 21.18; Pred. No. 0.057;
Matches 180; Conservative 131; Mismatches 324; Indels 219; Gaps 42;

QY 1 MKKKRVLLPLMALSTILVSTGNLEVIQAFVKQENRLNSESSESSQGLGYFSDLNFOA 60

DB 245 VQRGVILIGKDVAF---NGLSHPDVAKNIEQQKVISIEDSKPAKL-----ANVFAA 296
QY 61 PMVYTSSTGDISIPSELENIPESENOYFOGAIKSG-----FIKKKSEDTYF 108
DB 297 GNLYDVNTROVNNNTNKKRPITDTRKDNIAI--SGEAGSMYGRNIKFYTLDG----- 350
QY 109 ATSDNH--VTMWDDQEVINKASNSKIRLEKRLYQIKIQYORENTEKGLDFKLYWT 166
DB 351 --AGVNHQVIFAEDDILNITDDGNS-----RLNKVADYVR---VVGKDIEL--- 393
QY 167 DSQNKKEVISSDNLQPLBELKSSNSRKRKTSAGPYTPDDNDNGIPD--SLEVEGYVDV 225
DB 394 -----ANNGOIHADQQLINATGHVKLNGSSVISNNNGISALNITLENATVSA 443
QY 226 KNRFTLSPWISNHEKKGLTKRYKSSPEKWSSTADPYSD-----FEKYTGRIIDN 275
DB 444 NN-----LSFRYNTDKLNNLSKVS-----ARADLQSGNLNDKASVLAHKLLNLSND 493
QY 276 VSPARHPVLAAYPIVHVDMENIILSKNEDOSTONTSETRTISKNTS--TSRTHSEV 332
DB 494 VSLNQSRLSA-----NNLKIKKVRDLNLNSELSANNLTSMNITLKNKSKFT 544
QY 333 HGNAEVAHANTSTSTHTSEVHNAEVAHAVALDHSLSLAGEPTMAETMGLMTADTARLAN 392
DB 545 AGNMTLANTNNTVTLNDSLANNLTLNVTKNVTLNDASKS--ANKLDLNTDVTNLSK 603
QY 393 IRVYNTGAPRYNVLPTTSVLGKNQTLATIKAEQOLSQILAPNNYPSKNLAPIALNA 452
DB 604 -STLSAGLTFKKV-----KNVTL-----NNNSE--LAANN-----LSLNA 636
QY 453 QDDFSPTITMNYNOFLEKT--KQLRL--DTDOYGNATYTNFENGKRVDTGWSWS 507
DB 637 -----SHVYTLNNSKLSAOKADIKAVNLTLNDTE---LTANKLDINSTYTINNGTIA 687
QY 508 EVLPQIOETTARIIFNGKDLNVERRI--AAVNPSPLETTKPDMLKEALKIAGFNEPN 566
DB 688 GIFANI--TEKLNNKEKALLAEQNLNFTYNGSH--YENKGDYSKDAVTFSSKNS-- 741
QY 567 GNLOYGKDTTEDEFNPDQTSOIKNQLAEIENATNI-----YTVLDKIKLNAK----- 615
DB 742 -----DFTSGSKLVNAQNL--KVNNVNFITISOGDITLIGNTVLNASTFT 787
QY 616 -----MNLIDKRRHYDRNNIYAGAD---ESYVR-----EAKREVI 649
DB 788 NSGNLTIVTKLIDVGDIQFTNKGMLTVGEDLHISKTKITNDGKLISIKNLISSADFI 847
QY 650 NS-----STEGILLNIDKIRKILSGYIEVEDTEGLEKEVINDRYDM--LNIS 695
DB 848 NNGTILGTEALKIATKGNFTKKEKAI--LASNSLDSIVAECKKTFNNGTIESGKNLIT 905
QY 696 S---LRQDGKTFIDFKKYNDKPLIYSN-----PNKVVNVAATKENTIINPSENGDTS 746
DB 906 NTGAFNLVANDNATIRSFGLNITSTGCVSNNGTLLISNERLN--ITSANFTNES--NGTYM 961
QY 747 TNGIKKILIFSKG 760
DB 962 SNGLNIT--AKQG 973

Search completed: December 2, 2001, 13:54:52
Job time: 527 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:51:49 ; Search time 44.8 Seconds
(without alignments)
625.266 Million cell updates/sec

Title: US-09-747-521-4
Sequence: 1 MKRRKVLIPALSTILVS.....TSTNGIKILIFSKGVEIG 764

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3800	97.1	764	1 PAG_BACAN	P13423 bacillus an
2	235.5	6.0	192	1 YPA_BACAN	P13422 bacillus an
3	183	4.7	1803	1 JUL3_YEAST	P47024 saccharomyc
4	182	4.7	1658	1 YME7_YEAST	003661 saccharomyc
5	173.5	4.4	1230	1 SMC3_YEAST	P47037 saccharomyc
6	173.5	4.4	1271	1 Y338_MYCGE	P47380 mycoplasma
7	166.5	4.3	1928	1 MYSL_YEAST	P08964 saccharomyc
8	162	4.1	2116	1 MYSL_YEAST	P08799 dictyostell
9	162	4.1	2869	1 RBPI_PLAVB	000798 plasmodium
10	161	4.1	1957	1 YD86_SCHPO	010411 schizosacch
11	160	4.1	1790	1 US01_YEAST	P25386 saccharomyc
12	159.5	4.1	1630	1 MSPI_PLAFK	P04932 plasmodium
13	159.5	4.1	1639	1 MSPI_PLAFW	P04933 plasmodium
14	158.5	4.1	1487	1 MDS3_YEAST	P53094 saccharomyc
15	158	4.0	1385	1 PART_SCHPO	013735 schizosacch
16	157	4.0	1000	1 S155_YEAST	P43612 saccharomyc
17	156.5	4.0	803	1 SM16_YEAST	P09959 saccharomyc
18	156.5	4.0	1208	1 PCPI_SCHPO	Q92351 schizosacch
19	156.5	4.0	1251	1 RBP2_PLAVB	000799 plasmodium
20	156	4.0	1744	1 TANX_XENLA	001550 xenopus lae
21	155.5	4.0	1018	1 FNBA_STRAU	P14738 staphylococ
22	155.5	4.0	1024	1 R1P3_MOUSE	P97434 mus musculu
23	155.5	4.0	1358	1 SIR4_YEAST	P11978 saccharomyc
24	155.5	4.0	1541	1 IGAI_HAEIN	P42782 haemophilus
25	155.5	4.0	1545	1 IGAI_HAEIN	P45385 haemophilus
26	153	3.9	1570	1 P3K1_DICDI	P54673 dictyostell
27	153	3.9	1726	1 MSPI_PLAFW	P50495 plasmodium
28	151.5	3.9	2334	1 WARP_BACSU	007893 bacillus su
29	151	3.9	1218	1 MGPC_MTCPN	050341 mycoplasma
30	150.5	3.8	730	1 GLN3_YEAST	P18494 saccharomyc
31	150.5	3.8	1116	1 YK54_AQUAE	O67838 aquifex aeo
32	150	3.8	1420	1 SRB9_YEAST	P38931 saccharomyc
33	150	3.8	3418	1 BRCA2_HUMAN	P51567 homo sapien

ALIGNMENTS

RESULT	1	PAG_BACAN	STANDARD	PRT	764 AA.
AC	P13423				
DT	01-JAN-1990 (Rel. 13, Created)				
DT	01-JAN-1990 (Rel. 13, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	PROTECTIVE ANTIGEN PRECURSOR (PA) [CONTAINS: PA-20; PA-63].				
GN	PAG.				
OS	Bacillus anthracis.				
OG	Plasmid pXOI.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;				
OC	Bacillus/staphylococcus group; Bacillus.				
OX	NCBI_Taxid-1392;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MDLINE-69172073; PubMed-3148491;				
RA	Welkos S.L., Lowe J.R., Eden-McCutchan F., Vodkin M., Leppla S.H.,				
RA	Schmidt J.T.;				
RT	"Sequence and analysis of the DNA encoding protective antigen of				
RT	Bacillus anthracis.";				
RL	Gene 69:287-300(1988).				
RN	[2]				
RP	DOMAINS.				
RX	MDLINE-91332080; PubMed-1651334;				
RA	Singh Y., Kimpel K.R., Quinn C.P., Chaudhary V.K., Leppla S.H.;				
RT	"The carboxyl-terminal end of protective antigen is required for				
RT	receptor binding and anthrax toxin activity.";				
RL	J. Biol. Chem. 266:15493-15497(1991).				
RN	[3]				
RP	X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).				
RX	MDLINE-97192099; PubMed-9039918;				
RA	Petosa C., Collier R.J., Kimpel K.R., Leppla S.H., Liddington R.C.;				
RT	"Crystal structure of the anthrax toxin protective antigen.";				
RL	Nature 385:833-838(1997).				
CC	-1- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,				
CC	AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE				
CC	DEATH. PA IS THOUGHT TO BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC				
CC	CELLS, THEREBY FACILITATING THE INTERNALIZATION OF LF OR EF. PA				
CC	ASSOCIATED WITH LF CAUSES DEATH WHEN INJECTED, PA ASSOCIATED WITH				
CC	EF PRODUCES EDEMA. PA INDUCES IMMUNITY TO INJECTION WITH ANTHRAX.				
CC	-1- SUBUNIT: ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT PROTEINS, A				
CC	PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN EDEMA FACTOR				
CC	(EF). NONE OF THESE IS TOXIC BY ITSELF, ONCE ACTIVATED. PA FORMS				
CC	HEPTAMERS WHICH INSERT INTO MEMBRANES AND FORM CATION-SELECTIVE				
CC	CHANNELS.				
CC	-1- SUBCELLULAR LOCATION: SECRETED.				
CC	-1- DOMAIN: THE C-TERMINAL PART OF PA IS REQUIRED FOR RECEPTOR BINDING				
CC	AND TOXIC ACTIVITY.				
CC	-1- PTM: PROTEOLYTIC ACTIVATION BY FURIN CLEAVES THE PROTEIN INTO TWO				
CC	PARTS, PA-20 AND PA-63, THE LATER HEPTAMERIZE.				
CC	-1- SIMILARITY: TO C. PERRINGENS IOTA-B TOXIN AND TO VIP1 TOXINS IN				
CC	BACILLUS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M22589: AAA22637.1; -

DR PDB: 1ACC; 11-FEB-98.

DR InterPro: IPR003896; Binary_toxB.

KW Toxin; Plasmid; Calcium-binding; Signal; 3D-structure.

FT SIGNAL 1 29 PROTECTIVE ANTIGEN.

FT CHAIN 30 764 PA-20.

FT CHAIN 30 196 PA-63.

FT CHAIN 197 764 PA-63.

FT DOMAIN 30 287 DOMAIN 1, CALCIUM-BINDING.

FT DOMAIN 288 516 DOMAIN 2, HEPTAMERIZATION.

FT DOMAIN 517 764 DOMAIN 3.

FT CA_BIND 206 206

FT CA_BIND 208 208

FT CA_BIND 210 210

FT CA_BIND 217 217

FT CA_BIND 764 AA; 85811 MW; 84BB22690FEAAB5 CRC64;

SEQ SEQUENCE

Query Match 97.1%; Score 3800; DB 1; Length 764;

Best Local Similarity 97.6%; Pred. No. 2.7e-179; Mismatches 16; Indels 0; Gaps 0;

Matches 746; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 MKRRVLPILMALSTIIVSSTGNLEVIOAEVKEOENRLNESSSOGGLGGYFSDPLNFOA 60
 DB 1 MKRRVLPILMALSTIIVSSTGNLEVIOAEVKEOENRLNESSSOGGLGGYFSDPLNFOA 60
 QY 61 PMVVTSTTGDLSPSSLENIPISENQYFQSAIMSGFTKKKSDXYTATSADNHTVMV 120
 DB 61 PMVVTSTTGDLSPSSLENIPISENQYFQSAIMSGFTKKKSDXYTATSADNHTVMV 120
 QY 121 DDOEVYINANSNKRILEKGRLOYKIOYORENPEKGLDFLYTDSQNKKEVYSSDNL 180
 DB 121 DDOEVYINANSNKRILEKGRLOYKIOYORENPEKGLDFLYTDSQNKKEVYSSDNL 180
 QY 181 QLPPELKOKSSNRKRSTASAGPTVPDRDNDGIPDSLEVEGYTVDVKNKRTFLSPISNTH 240
 DB 181 QLPPELKOKSSNRKRSTASAGPTVPDRDNDGIPDSLEVEGYTVDVKNKRTFLSPISNTH 240
 QY 241 EKKGLTKKSSPEKWSIASDPYSDFEYVGRIDKNVSEARHPPLVAAVPIVHVDMENTIL 300
 DB 241 EKKGLTKKSSPEKWSIASDPYSDFEYVGRIDKNVSEARHPPLVAAVPIVHVDMENTIL 300
 QY 301 SKNEDOSTONTDETRTISKNTSTSRHTSEVHGAEEVHAHNTSRKRTTSEVHGAEEVHA 360
 DB 301 SKNEDOSTONTDETRTISKNTSTSRHTSEVHGAEEVHAHNTSRKRTTSEVHGAEEVHA 360
 QY 361 VAIDHSISLAGEERTWAETMGINTADTARLANIRVNTGTAPYVNLPTTSVLGKNOTL 420
 DB 361 VAIDHSISLAGEERTWAETMGINTADTARLANIRVNTGTAPYVNLPTTSVLGKNOTL 420
 QY 421 ATIRAKENQLSQIILAPNNYPSKNLAPIALNAQDDFSSPTTMANYNOFLDEKTKQRLD 480
 DB 421 ATIRAKENQLSQIILAPNNYPSKNLAPIALNAQDDFSSPTTMANYNOFLDEKTKQRLD 480
 QY 481 TDQVYGNATYATNFENGRRVDTGSGNWSSEVLPOIOETTARIIFNGDNLVERRIAAVNP 540
 DB 481 TDQVYGNATYATNFENGRRVDTGSGNWSSEVLPOIOETTARIIFNGDNLVERRIAAVNP 540
 QY 541 DPLETTKPDMLLEKALTAIFGNEPENGMLYOAGKDITFEFDFPDQOTSQNTKNOIAELNA 600
 DB 541 DPLETTKPDMLLEKALTAIFGNEPENGMLYOAGKDITFEFDFPDQOTSQNTKNOIAELNA 600
 QY 601 TNITVYLDKIKLNAKMNLIDKRFHYDRNNIYAVGADESYYKKAHREYINSTEGLILNI 660
 DB 601 TNITVYLDKIKLNAKMNLIDKRFHYDRNNIYAVGADESYYKKAHREYINSTEGLILNI 660
 QY 661 DKDIRKILSGYIVLEDEEGLEKVIYNDRIYDMLNITSLRQDKTFIDPFKKYDKPLIYISN 720

DB 661 DKDIRKILSGYIVLEDEEGLEKVIYNDRIYDMLNITSLRQDKTFIDPFKKYDKPLIYISN 720
 QY 721 PMYKYNVAVTKENTINPSENGDSTNGIKKILFSKKGYEIG 764
 DB 721 PMYKYNVAVTKENTINPSENGDSTNGIKKILFSKKGYEIG 764

RESULT 2

YPA_BACAN STANDARD; PRT: 192 AA.

ID YPA_BACAN

AC P13422;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 01-NOV-1990 (Rel. 16, Last annotation update)

DE HYPOTHETICAL 21.6 KDA PROTEIN IN PROTECTIVE ANTIGEN 5'REGION.

OS Bacillus anthracis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=1392;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=89172073; PubMed=3148491;

RA Welkos S.L., Lowe J.R., Eden-McCutchan F., Vodkin M., Leppla S.H.,

RA Schmidt J.J.;

RT "Sequence and analysis of the DNA encoding protective antigen of

RT Bacillus anthracis."

RL Gene 69:287-300(1988).

-1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M22589; AAA22636.1; -

DR HSSP: P13423; IACC.

KW Hypothetical protein; Transmembrane.

FT TRANSMEM 162 182 POTENTIAL.

SEQ SEQUENCE 192 AA; 21620 MW; DBC7150AE78F8AFA CRC64;

Query Match 6.0%; Score 235.5; DB 1; Length 192;

Best Local Similarity 34.6%; Pred. No. 1.8e-05; Mismatches 47; Indels 23; Gaps 6;

Matches 56; Conservative 36; Mismatches 47; Indels 23; Gaps 6;

QY 616 MNILIRDRFHYDRNNIYAVGADESYYKKAHREYINSTEGLILNIDKIRKILSGYIEI 675

DB 1 MNILIRDRFHYDRNNIYAVGADESYYKKAHREYINSTEGLILNIDKIRKILSGYIEI 675

QY 676 EDTE-----GLKEVYNDRIYDMLNITSLRQDKTFIDPFKKYDKPLIYISPN 722

DB 60 KPSNHLNTPVYITLAKDGSVGLYRLVLS-----DGAQFLDNKDKEDMWSRLV-DPG 112

QY 723 YKVNYYAVTKEN-TIINPSENGDSTNGIKKILFSKKGYEI 763

DB 113 DDVYVYAVTKEDENAVTRDENGNIA-NKLKNTLVLSGKIKEI 153

RESULT 3

YJL3_YEAST

ID YJL3_YEAST

AC P47024; P87192;

DT 01-FEB-1996 (Rel. 33, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE TRANSPOSON TY4 207.7 KDA HYPOTHETICAL PROTEIN.

GN TY4B OR YJL113W OR J0780.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1679;
 RX MEDLINE=97103775; PubMed=8948101;
 RA Czaplich C., Korde E., Pujol A., Jauniaux J.-C.;
 RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
 RT SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, CCT7, GEF3, two tRNA genes,
 RT three remnant delta elements and a Ty4 transposon";
 RL Yeast 12:1471-1474(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 249389; CAAB9409.1; -
 DR SGD: S0003649; YOL113W.
 DR InterPro: IPR001584; Rve.
 DR InterPro: IPR001878; znf_CCNC.
 DR Pfam: PF00665; rve; 1.
 DR SMART: SM00343; znf_C2HC; 1.
 KW Transposable element; Hypothetical protein.
 SQ SEQUENCE 1803 AA; 207693 MW; 16DCD7284ABD52D3 CRC64;

Query Match 4.7%; Score 183; DB 1; Length 1803;
 Best Local Similarity 20.1%; Pred. No. 0.14;
 Matches 167; Conservative 117; Mismatches 283; Indels 264; Gaps 40;

OY 25 EVIOAEVKQENRLNLESSESSOGLGYPSDLNFOAPMVYVTSSTGSLSPSELSEINIPS 84
 DB 1073 ELIQEBLKTN---HETSPKESIG---TNVFRNTNNEISLKTGDTSLPITLESINN 1126
 OY 85 ENQYFOSAIWSGFIKVKSDERTFATSADNHVTWVWDDEVINKASNSKIRLEKGRLYO 144
 DB 1127 HH-----SNDY-----STNKVE----- 1138
 OY 145 IKIOYOREN---PTEKGLDFKLYMTDSQKKKEVISDNQLPELKOKSSNSKKRSTSA 200
 DB 1139 ---KFEKENHHPREDIDYDM---SDQDMESNCQDGNLKLKYTDKNVPTDNGTNY 1190
 OY 201 GPTVPDRNDGIPDLSLEVGTYVDVKNKRTFLSPWISNIHEKK-----GLTRY--- 248
 DB 1191 SPR-----LEQNEASGSPYQYVKNKSAFLNKEFSSSLMKRRKRHRDKNNLSLTGELE 1242
 OY 249 -----KSSPEKMTASDPYSDFEKYVGRIDKNVSPEARHPLVAAYPIVHVDMEN 297
 DB 1243 RDKRSKKNRVKLIIPDNMTETVSAPIRAYUENAIKSNPDLEKHEVKAQY---HKEION 1299
 OY 298 ILSKREDSQNTQDSEETISKN--TSPSRITSEVHG--NAEVANISTSTHSEVH 353
 DB 1300 LQDMKVFVDVVKYSRSE---IPDNLIVPNTJFTKRRNGIKYKARIVCRGDTSPDITYSVI 1356
 OY 354 GNAEVAIVAVIIDSLSLAGERT--WAETMGLTADT-ARLANINRYVTGTAIVNVP--TT 410
 DB 1357 TTESLNHNHIKIFLMTANRRNFMKTLIDINHAFLYAKLEE-----IYIPRHR 1406
 OY 411 SLVLGNQTLATIKAKENQ-----LSQI-LAPNNYYP-----SKNLAPIALNAQDD 455
 DB 1407 RCVVKLNKALYGLKQSPKEMNDHLROYLNGIGLKDNSYFGLYOTEDKNLM-LAVYVDDC 1465
 OY 456 FSGSTPTMYNOFL-ELEKTQOLR-----LQTDQYGNATATNEENGRIV----- 498
 DB 1466 VLAASNEQRLDEIFINKLSKSFELKTGTLLDLDVLDTD-LIGMDLVYNNKRGITDLTLKSF 1524
 OY 499 --RVDTGSNMSEVLPQIOETTARIIFNGKDLNVERRIAANVP--SPLETTKPD-----M 550
 DB 1525 INRMD--KKRYNEBLKIRKSSIPHMSTYK-----IDPKKDVLQWSEEFROGYVL 1571
 OY 551 TLKEALKIAGFNEPNCNLOYOQKIDITEFPNPDQOTSQINKQALAEANTNTYTLDKI 610
 DB 1572 KIQOOL-----GELYV--RHRCRYDIEFAVVKVARLVYPEREFYMYTKIIQYL 1620
 OY 611 KLNAMKMLIRDRFHFDRN-----NIAGVADSVEYKARE-----VI----- 649
 DB 1621 -----VRKIDGIRHRCNCKNDKVIATIDASVSGSEYDAQSRIGVITLYGMNIFNVIS 1673
 OY 650 NSSTEGILLNIDKDIRKILSGY-----IWEIEDT-----EGIK-- 682
 DB 1674 NKSTNRCVSTAEALHAYEGVADSETLKYTLKELGCGNDIVMTDTSKPAIOGLNRSY 1733
 OY 683 -----EYINDRYMLNISSLRQDGKFI-----DFKRY 710
 DB 1734 QQPKRFYWKITETIKERIKESIKILKTIGKNIDLTKRPVSASDFKRF 1784

RESULT 4
 ID YM67_YEAST STANDARD; PRT; 1658 AA.
 AC 003661; Q04988;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHETICAL 187.1 KDA PROTEIN IN GUAL-ERG8 INTERGENIC REGION.
 GN YMR219W OR YMR261.13 OR YMR959.01.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE OF 1-711 FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
 RL Walsh S.V.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 608-1648 FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A.,
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 249809; CAAB9934.1; -
 DR EMBL: 249939; CAAB9190.1; -
 DR SGD: S0004832; YMR219W.
 KW Hypothetical protein.
 SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;

Query Match 4.7%; Score 182; DB 1; Length 1658;
 Best Local Similarity 19.3%; Pred. No. 0.14;
 Matches 161; Conservative 133; Mismatches 308; Indels 234; Gaps 39;

OY 25 EVIOAEVKQENRLNLESE-SSOGLGLGYPSDLNFOAPMVYVTSSTGSLSPSELSEINIP 83
 DB 287 EGYEMLEDDIDVDESQKESQGAECTEHS-VDFSXYMQPRTDNTKIPYIEKYESDHE 345
 OY 84 SENQYFOSAIWS-GFIKVKASDE-----YTFATSADN---HYTMVWDQEVYINKASN 131
 DB 346 VHOYSEDAQFADGVSNIIVDSEDESOAESYSANAENVYHNEHLEDDKELIEDIES 405
 OY 132 SNKIRLEKGLYQIKIOYORENPTKEGLDKFL-----YWTDSQNK 171

```
Db 406 SDS-----ESQSAQESQGSIDPEFYKMKNEKSTSETEMTSESROGFAKDAVTK 456
Qy 172 KEVTSNDLQPEL-----KOKSSNRKRKSTAGPTVPDRDNDGIPDSL 216
Db 457 NKVEQGEDEBEKDDITRSILDKNFHGNKNSKSEYSENVLENEDPAIVERNO_ IND- 513
Qy 217 EVEGYTVYVKKRFFL---SPWISNIHF-KKGLTKYSS-----PEKWTASADPY---S 263
Db 514 -VEGYDVYKGSVESDLHESP--DNLXDLAARAMQFOQSRRNSNCPQKEQVSESYLGS 570
Qy 264 DFEKATGR-IDKNVSPGARHPVLAAYPIVHVDMENILSKNEDOSTQJTOSETRTISKNT 322
Db 571 NGSNLSGRSLDES---EQQILPKD-----FTGEENNNNLKTGDGSSSVEI 613
Qy 323 STSRHTSEVGNAAVEH-ANTSTSRHTSEVGNAAVEHVAIDHSLAGERTAEWGL 381
Db 614 EKEVYSEKKLGGSTREKELVPLSTDTTINNSSLGNEDSIYSLDADASENLDTVPLEI 673
Qy 382 NTA-----DTARLANIRIVNTGAPLIYVNLPTTSLVGNKQTLAT 422
Db 674 KTTPEYEVVISSEYSSSTSYEDNTVAMPQVEY---TSPPMND-PFNSL-----ND 721
Qy 423 IKAKENOLSOILAPNNYPSKNLAPIALNADDFSPPTIT-----NYNOFLELEKT 474
Db 722 YEKHDLKSLTAA-----LAPATFKDAEVEVAGVTSCLSLSTSGHNITHTSKET 773
Qy 475 KQLRLDQVYGNATATYVFNENGRRVVDGSMSEVLPOIQTETARIIFNGDLNVERRI 534
Db 774 KOVS-DLDEESTENTYFENENTG---DENKNQSKNFPGVANSTDKSTEDND---EKYF 824
Qy 535 AAVPSPDLETTKPDMLTKEALTAFGFENPENGNYQYQKITERDFNF-DOQTSQN--- 590
Db 825 SAINTN---VTGDSCEDIETASVNEE---NLRYCEKMNEMKSSGDECVKQNDG 877
Qy 591 IKNOA-----ELMATNITYVLDKIKLNKMNILIRDKRFRHNNIAVGADESUYKE 643
Db 878 SKTQISFTSDPNDPQESNDMTFFSSTK-----YKVRNSDLEDESLAKE 922
Qy 644 AHR-EVIN-----STEGELLNIDKIDIRKILSGYIEIEDTEGLK 682
Db 923 LTKAEVVDKLDDESEDSYEDDYADPEPGNDEGSNENIVKGTGK-----DTLGIV 972
Qy 683 EVIDRVDYMLNLSRQGGKTFIDEKKYNDKPLXISNPNKKVNVYATKENTIIIN 738
Db 973 EPENEKVN-----KYHEEETLEFANVSSVNVQNKMDHTDVIIN 1010

RESULT 5
SMC3_YEAST STANDARD: PRT: 1230 AA.
AC P47037;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CHROMOSOME SEGREGATION PROTEIN SMC3 (DA-BOX PROTEIN SMC3).
GN SMC3 OR YJL074C OR J1049.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W303;
RX MEDLINE=97474309; PubMed=9335333;
RA Michaelis C., Cloak R., Nasmyth K.;
RT "Coheins: chromosomal proteins that prevent premature separation of
RT sister chromatids.";
RL Cell 91:35-45(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Rose M., Koelter P., Entian K.D.;
RL Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.
RN [3]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Sor F.J.;
RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y14278; CA74655.1; -
DR EMBL: Z49349; CA89366.1; -
DR EMBL: X88851; CA61313.1; -
DR SGD: S0003610; SMC3.
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02483; SMC_C; 1.
DR Pfam: PF02463; SMC_N; 1.
DR Mitos: ATP-binding; Coiled coil; Nuclear protein.
FT NP_BIND 32 39
FT DOMAIN 172 482
FT DOMAIN 685 1041
FT COILED COIL (POTENTIAL).
FT COILED COIL (POTENTIAL).
SQ SEQUENCE 1230 AA; 141336 MW; B152D88F7780341F CRC64;

Query Match 4.48; Score 173.5; DB 1; Length 1230;
Best Local Similarity 20.04; Pred. No. 0.24; Indels 247; Gaps 39;
Matches 186; Conservative 137; Mismatches 358;

Qy 22 GNLEYIAQAEVQENRLNSESSESSGGLGYFSDLNFOAPWVY--SSTTGDISPSEL 79
Db 201 GELNSKLSMEMQERKELEYMLENRKRIYFTLYDRLEINVOMERLDDGYN----- 254
Qy 80 ENISENOYFO-----SAWSGPIKVKKSDPYFPASADNHVTWVDDQ 123
Db 255 NTVYSSSEQYIQLDKREDMDIDOVSKLSSI-EASLIKINATDLOAKLRESEISOKLTNV 313
Qy 124 EVINK-----ASNSNKRLEKGRLOYIK-IQYQRENTPEKGLOPKLWTDSQKKVEIS 176
Db 314 NVKIKDVQOOIESNEEQNLNDSATLKEIKTSIIEOKKOLSKILPRYOELT---KEEMAY 369
Qy 177 SDNLQPELKOKSSNRKRKSTAGPTVPDRDNDGIPDSLEVEGYTVYVKKRRTFLSPWI 236
Db 370 --KQLASLQCKQNDLILKKEVYARFKSKDERDWMHSEIE-----ELKSS-----I 414
Qy 237 SNIEH-KKGLTKYSSPEKWTASADPYSDPEKVTGRIDKNVSPGARHPVLAAYPIVHVD 295
Db 415 QNLNELESQLOMDRSTLKKQYSAID-----EIEELLDISINGPPTKCOL-----EDFDS 463
Qy 296 ENILSKNEDOSTQJTOSETRTISKNTSTSRHTSEVGNAAVEHVAIDHSLAGERTAEWGL 355
Db 464 ELIHLKQKLSLDTRKLMRKEQKLQTVLETLSDVQON-QRVNENETMS-----LAN 517
Qy 356 AEVHVAIDHSLAGE-----RTWATWG-----LNTADTARLANA 391
Db 518 GIINWKEITELKLSPEVFTGLDELIVKNDKYTCACAVIGNSLFIHVDTEETATILM 577
Qy 392 NIRVYNTGAPLIYVNLPTTSLVGNKQTLATIKAKENOLSOILAPNNYPSKNLAPIALN 451
Db 578 NELYRMKCGRYTF--IPLNRLSDSDVAFPSNTTQIOFTPLIKIKIKKEPFEKA----- 630
Qy 452 AODDESSPTIMNNYQFLELEKTKQLR--LDTPQV-IGNLAT--YNNENGRVAVDTGSN 505
```



```

Db 631 VKHVFQKTVVKGDLGGGLAKKHKLNATITLDDRADKRGVLITGGLDQHKRTRESLKN 690
QY 506 MSEVLQIOETARTIENGKDLVERRIAANPS-----DPLETT 546
Db 691 LNSRSQHKILEELDFVRNELNDITKIDVGNKIKVSNDSRESVLTINIEVYRTSLNTK 750
QY 547 KPD-MTLKALK-IAGFNEPNCNLQYOGKDIPEFEN-----PDQOSONIKNOLA----- 596
Db 751 KNEKLLESNLNIIILKELKLNTRTFADQKLMFTFENDLLQGEDSELSKEKRELESLTK 810
QY 597 -----ELNAIN-----IYTVLDKIKLAKKNMILRDKRFHYDRNNIAVG----- 635
Db 811 EISAHNKINTISDALEGITTTID--SLNAELESKLIPQNDLESKSEGDAFIFGLQD 868
QY 636 -----ADESVYKEAHREYINSS-----EGLL--N 659
Db 869 ELKEQLKESEVKEHONAVLEIGTVORETESLIAEETNNKILEKANNQOIRLLKLDN 928
QY 660 IDKDIRKISGYIEIEDTEGLKE-----VINDYDM-----LN-----I 694
Db 929 FOKSVKTKMTKTTLYTRRELQORIREIGLPELDALVNFSDITSQOLLRLDMMTEI 988
QY 695 SSLRQDGK-TFIDFKKYNDK-----LPLYSNPKYKNVYAVTKENT 735
Db 989 SGLKNVKNKRAFEKFKNERKDLARASELDESKDSIODLIYKQKQVNAV-----DST 1044
QY 736 IIMPSENDT-----STNGIKKILIFSK 758
Db 1045 FOKVSENFEAVFERLVRPGTKAHLIRK 1072

RESULT 6
Y338_MYCGE
ID Y338_MYCGE STANDARD: PRT: 1271 AA.
AC P47580;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYDROTHERICAL LIPOPROTEIN MG338 PRECURSOR.
GN MG338.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=756993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Frittmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bort K.F., Hu P.-C., Lueder T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 1023-1114 FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC ANCHOR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE MG307 / MG309 / MG338 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U39715; AAC71563.1; -.
DR EMBL: U01809; AAD12341.1; -.
DR TIGR: MG338; -.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Lipoprotein; Membrane; Signal;
KW Complete proteome.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 1271 HYPOTHETICAL LIPOPROTEIN MG338.
FT LIPID 27 27 N-ACYL DIGLYCERIDE (POTENTIAL).
FT SEQUENCE 1271 AA; 142492 MW; FCE6042067310A70 CRC64;
SQ

Query Match 4.4%; Score 173.5; DB 1; Length 1271;
Best Local Similarity 18.4%; Pred. No. 0.25;
Matches 181; Conservative 143; Mismatches 337; Indels 321; Gaps 42;

QY 14 STIVSSTGNLEVIQ-----AEYKQENRLNE-----SESSQGLGYFFSDNFGAP 61
Db 317 SKVLISKNLISLVKTYNLNSAVIDQYHYLLNKTELTTTITTTTGGTSSNLPDLK 376
QY 62 MVTYS--TTGDLISPSSELENIPSENOYFQSAIWGEIKYKKSDEYTFATSDNHYTMV 120
Db 377 FIKSSSATVYMKMSAMTKSQEVTSDNNGFN--VKSEFLKINPSSSGSDMSNTQSEFWK 434
QY 121 DDQEVINKASNS--NKIRLEKG-----RLYQIKIOYQENPTEKGLDKLY 164
Db 435 QVQALNNSQATATTFDAVRMESNSSQAQVYTSNLVLSKTKQKQKQKPYVVRGDAIY 494
QY 165 -----WDSQKKEVISSDNLQPELKQKSSNRKRSATGAPVPPDNDNGIP 213
Db 495 AFHIDGNYPLENSSPKNRKFEKQAEVLLMRFLQGTNNFSK-----DNVS 540
QY 214 DSLEVEGYTVDK--NKRITFLSPWI-----SNH 240
Db 541 FSVYLFPSNSEFRMANRNTTKLYTALTMLNGTSNNQKQVCDLAKKLNNTLS 600
QY 241 E-----KGLTKYKSSPKKSTASDPYSDEFKVTGRDK-----NVSPE 279
Db 601 ETIKQODFNSSLSQIKSSSESYIKAAKNLNFPTDLANIEQLQAVDRANNYIKLQKE 660
QY 280 ARHPLY-----AAYPIVHDMENILSKNEDOSTNT-----DSE 314
Db 661 AKESSTIGWQPLPYKRANDQSYSLAKFPNN-----NSDQSAOTLTKTTAAITSDNE 715
QY 315 TRTSKNTSTSRHTSEVHGNAEVHANTSTSRHTSEVHG-----NAEYHAVAIDHSLS 368
Db 716 P-FTEKKNQTLKLTTEVENKAKKELVEKKATYVSSSOYSLITLKSSQLDNDLDDLLLS 774
QY 369 LAGERTVAETWGLNTADTARLANIRYVNT-----GTAPIYVLPPTSLV 413
Db 775 ILTD-----SGIRGTIVANIFKMWYFKNTSFNNEDTSNKEKLGCFSPFNDLVKQALY 828
QY 414 LKNQTLATIKAKENQSLIAPNNYP--SKNLALIALNAQDDFSPTITMNYNQL- 469
Db 829 IRSMQNLV--SKER-----FGYKKGDSVNSTSTMOMQKKAQSHSTSSVNOJTL 877
QY 470 -----ELEKTKQLRLDQYVGNI--ATNFNGRVRVDGSMSEVLQIOETT 517
Db 878 DLAKKAEKELEDPNQ--DAEYKYMRFLQALMLVANG-----AQNYKNLLOALPIG 928
QY 518 ARIIFN--GKDLNVERRIAANVSDPLETTKEDMTLKEALKITAGF-----ENEGCN 568
Db 929 TRAFVSMVTGYDKN-----PSATVYQKTKSTSSANENPNRFLQN 969
QY 569 LQY-QGNDITEFPNPOQOSONIKNOLAELANITYTVLDK-----IKNAKNMILRD 622
Db 970 PNTQGESEIMFN--DKQPTIOPDSL--LESNTYRTFDEPNNSVALSNKQO-GSSD 1022
QY 623 KRPHYDRNNIYAGDESVKEAHREYINSSTEGGLINDKIRKIL-----SGYIETED 677

```

```

Db      1023 KKFFGFGNGLTITNSQ-----ISFASGLTQQQLFNNGOLITATDRAKALSQYKD 1073
Oy      678  TEGLEKIVIND-----RYDMLNITSLEODGKTIFD---RK----- 709
         |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      1074 KFTLLSINKTSDDAELNAFCGLLHRSVNVDTNNLSRFNSREDELPIDDNKKFLLNVVD 1133
Oy      710 -----YNOKLPLXI-----SNPNYK-VNYVAATKEVTILNP----- 740
         |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      1134 RLDDLFHFHFECGYVLAKRANTDGSNTGECYKTVGTVPFLTNDSIIQTLSYVIQITNDVVK 1193
Oy      741 -----ENGDTSTNGIKIILIFS 757
         :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      1194 ISSNMKNKANSGLKKEIFMS 1215

RESULT_7
MYSL_YEAST STANDARD: PRF: 1928 AA.
AC P089564:
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN-1 ISOFORM (TYPE II MYOSIN).
GN MYO1 OR YHR03FM.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycas.
OX NCBI_taxid=4932;
RN RP SEQUENCE FROM N.A.
RC STRAIN-S288C.
RX MEDLINE=91088308; PubMed=2263482;
RA Sweeney E.P., Watts F.Z., Pocklington M.J., Orr E.;
RT "The MYO1 gene from Saccharomyces cerevisiae: its complete nucleotide
RL sequence."
RN Nucleic Acids Res. 18:7147-7147(1990).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favellio A., Fulton L., Gatlton S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jter M., Johnston L., Langston Y.,
RA Laclelle P., Louis E.V., Macri C., Mardis E., Meneses S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RN VIII.".
RL Science 265:2077-2082(1994).
[3]
RN SEQUENCE OF 1-760 FROM N.A.
RP STRAIN-S288C;
RC MEDLINE=88111539; PubMed=3322809;
RA Watts F.Z., Shiels G., Orr E.;
RT "The yeast MYO1 gene encoding a myosin-like protein required for cell
RL division.";
RN EMBO J. 6:3499-3505(1987).
CC -!- FUNCTION: REQUIRED FOR CELL DIVISION.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X53947; CA37894.1; -
DR EMBL: X06187; CA29550.1; -
DR EMBL: U10399; AAB68872.1; -

```

DR	PIR: S05806; S05806.
DR	PIR: S12323; S12323.
DR	PIR: S46773; S46773.
DR	HSSP: P08799; 1MND.
DR	SGD: S001065; MYO1.
DR	InterPro: IPR000048; IQ.
DR	InterPro: IPR001609; myosin_head.
DR	Pfam: Pf00063; myosin_head; 1.
DR	PRINTS: PD00193; MYOSINHEAV.
DR	ProDom: PD000355; myosin_head; 1.
DR	SMART: SM00015; IQ; 1.
DR	SMART: SM00242; MYSC; 1.
DR	PROSITE: PSS0096; IQ; 1.
DR	Protein: Actin-binding; APP-binding.
KW	MYOIN
FT	DOMAIN 1 793
FT	DOMAIN 794 823
FT	DOMAIN 856 1911
FT	NP_BIND 180 187
FT	DOMAIN 460 529
FT	MOD_RES 703 703
FT	CONFLICT 36 36
FT	CONFLICT 46 46
FT	CONFLICT 59 59
FT	CONFLICT 86 86
FT	CONFLICT 330 330
FT	CONFLICT 343 343
FT	CONFLICT 421 426
FT	CONFLICT 515 515
FT	CONFLICT 529 535
FT	CONFLICT 541 541
FT	CONFLICT 550 551
FT	CONFLICT 573 573
FT	CONFLICT 582 582
FT	CONFLICT 588 599
FT	CONFLICT 599 599
FT	CONFLICT 627 632
FT	CONFLICT 695 695
FT	CONFLICT 736 742
FT	CONFLICT 756 756
FT	CONFLICT 773 784
FT	CONFLICT 793 794
FT	CONFLICT 896 896
FT	CONFLICT 900 900
FT	CONFLICT 906 906
FT	CONFLICT 911 911
FT	CONFLICT 915 930
FT	CONFLICT 934 939
FT	CONFLICT 951 953
FT	CONFLICT 955 958
FT	CONFLICT 1002 1002
FT	CONFLICT 1049 1049
FT	CONFLICT 1056 1056
FT	CONFLICT 1060 1060
FT	CONFLICT 1085 1085
FT	CONFLICT 1123 1123
FT	CONFLICT 1133 1133
FT	CONFLICT 1144 1146
FT	CONFLICT 1159 1168
FT	CONFLICT 1179 1181
FT	CONFLICT 1184 1185
FT	CONFLICT 1188 1204
FT	CONFLICT 1224 1224
FT	CONFLICT 1228 1228
FT	CONFLICT 1253 1253
FT	CONFLICT 1311 1323
FT	CONFLICT 1400 1400
FT	CONFLICT 1454 1554

```

FT CONFLICT 1568 1568 D -> V (IN REF. 1).
FT CONFLICT 1630 1646 DLKQDHTTKVEMLN -> SEARSLKSGGND (IN
FT CONFLICT 1698 1704 MISSING (IN REF. 1).
FT CONFLICT 1725 1737 TLOLOMEONSRNG -> NTAANGTKREW (IN REF.
FT CONFLICT 1754 1757 FDDE -> LMM (IN REF. 1).
FT CONFLICT 1777 1777 D -> E (IN REF. 1).
FT CONFLICT 1788 1788 R -> T (IN REF. 1).
FT CONFLICT 1825 1825 S -> D (IN REF. 1).
FT CONFLICT 1882 1882 S -> W (IN REF. 1).
FT CONFLICT 1902 1904 FWR -> NSGKRDLADL (IN REF. 1).
SQ SEQUENCE 1928 AA; 223634 MW; 6F54C7611F43DC9F CRC64;

Query Match 4.38; Score 166.5; DB 1; Length 1928;
Best Local Similarity 19.28; Pred. No. 0.97;
Matches 162; Conservative 145; Mismatches 291; Indels 247; Gaps 39;

QY 20 STGNLEVIOAEVROENRLNSES-----SSQGLGYSFSDNLFQAPMVVTSSTGDL 73
DB 894 TVNELETTQDLNOEKENLNKKNESLNRKVTSTTLOKQFDL-----VSEKDEIS 944
QY 74 IPSEL-ENIPSENOYFQSAIWGFIKKKSDEYTFAT--SADNHYVMWVDDQEV-INKA 129
DB 945 REKLEVAQNEEAHQKIQ-----GLQETIRERENTLEKLSKNNELKQISDLNCDISKE 999
QY 130 SNS-----NKIRLE-----KGRLOYIKIQVORENTEKGLDFKLYWTD----- 167
DB 1000 QSSOSLIKESKLEKLEIKRLKDVINSKEEIKSFNDKLSSESDIDIKITVLEKKNCIA 1059
QY 168 -----SON-----KKEVSSDN-----LQPELKQSSNRKRSSTAGPT 203
DB 1060 MSRLQSLVTNSDLSRNSNENFKKAKALNQLNKKESLKKMKIDNKKELATFS--- 1116
QY 204 VPDNDNGIDPSLEVES-YTVYKNNKRTFLSPWISNTHKGLKLYKKSPEKSTASDPY 262
DB 1117 -KQRD-----DAVSEHKITAELEKETRI-----QLTEKYSNOK--IKREY 1154
QY 263 SDEPKYGRDKNVSPARHPVAAYIVIVDMENILSKNEDOSTONTSETTISKNT 322
DB 1155 SNFORET---KEOEKRRKSLV-----ESLDSKIKELEAR-----LQOEI 1192
QY 323 STRTHTSEVHGAEVANSTSRHTS-----EVHGAENVH 359
DB 1193 SLMOYLNRKISGNS-VETINISSTRSTSYSDDPDKEDIIKKYYDLOLAFETIRNLENE 1251
QY 360 AVAIDHSLAGERTMAET-MGLNTADTARLMANIRYVNTGTADIVNLPPTSLVLGNQ 418
DB 1252 ---IEEKNLISRLRFETRLASSFEDQKIKAQMKKIKLQIDMOPISIPDSIL---NE 1305
QY 419 TLATIRAKENQSOILAPNNYIPSKNLAIALNAQDFFSPITMANYNOELEKTOQLR 478
DB 1306 PLNCPKEDSKIDINKMLEVYL--KROLDIETRAHYD-AENALISLSKRFKID--GESS 1360
QY 479 LPTDOVGNATATVNFENGRVVDGSMWSEVLPQIET---TARIINGKOLNIVERRIA 535
DB 1361 LSSSDIT---KLKFESEERKVSLEDKLTKMPLRDRTNLPVGDIIKRRDISKYEERIR 1416
QY 536 AVNPSDELTTKPDMTLEALKIAFGNEPNCNLOY----- 571
DB 1417 YRK---LENYK---LQELT-----NESNGKLSQLTLDLRQSKSEALISEOLDRLQK 1462
QY 572 -----QKQITEPDEFNFDOOTSONIKNOL-----AELNATNIYTVLDK 609
DB 1463 DLESTERQKELSLSTTIKQKQOFENCMDDLOGNELRLREHIALKQAEEDKNNASIIK 1522
QY 610 IKL-NAKMNILIRDKRFHYORNNIATVAGADSVKVEAHREVINSSTEGLLINDIDIRKL 668
DB 1523 LKTONKKEKILNFEREM--ERNDSMDLOLETLE-----LKRVDVAKKIL 1565
QY 669 GYIIVE-----IEDTEGLKEVINDRYDMNLISSLRODCKTFIDFKKYNKDLPLYISNP 721

```

```

DB 1566 SDDLHLKLERLSAVEDRSQYTDIEINRLKEELN-CSLKAETNLKKEFATLKYKLETTNDS 1624
QY 722 NYKNV 726
DB 1625 EAKIS 1629

```

```

RESULT 8
MYS2_DICDI STANDARD; PRT: 2116 AA.
AC P08799;
ID MYSL2_DICDI STANDARD; PRT: 2116 AA.
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN II HEAVY CHAIN, NON MUSCLE.
GN MHCA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.
OX NCBI_Taxid=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87092266; Pubmed=3540939;
RA Marick H.M., de Lozanne A., Leinwand L.A., Spudich J.A.;
RT "Conserved protein domains in a myosin heavy chain gene from
RT Dictyostelium discoideum.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RX STRAIN-AX2;
RX MEDLINE=90353583; Pubmed=2387408;
RA Lueck-Vielmeier D., Schleicher M., Grabatin B., Wippler J.,
RA Gerisch G.;
RT "Replacement of threonine residues by serine and alanine in a
RT phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
RL FEBS Lett. 269:239-243(1990).
RN [3]
RP PHOSPHORYLATION SITES.
RX MEDLINE=88112226; Pubmed=2828113;
RA Nagle G., Noegel A., Scheel J., Gerisch G.;
RT "Phosphorylation of threonine residues on cloned fragments of the
RT Dictyostelium myosin heavy chain.";
RL FEBS Lett. 227:71-75(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX MEDLINE=95345066; Pubmed=7619795;
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutcliffe K., Holden H.M.,
RA Rayment I.;
RT "X-ray structures of the myosin motor domain of Dictyostelium
RT discoidium complexed with MgADP, BeFx and MgADP.ALFA-";
RL Biochemistry 34:8960-8972(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX MEDLINE=95345067; Pubmed=7619796;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
RT truncated head of Dictyostelium discoidium myosin to 2.7-A
RT resolution.";
RL Biochemistry 34:8973-8981(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=96206189; Pubmed=8611530;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II).ADP.vanadate complex of the
RT Dictyostelium discoidium myosin motor domain to 1.9-A resolution.";
RL Biochemistry 35:5404-5417(1996).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX MEDLINE=97452580; Pubmed=9305951;
RA Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT "X-ray structures of the MgADP, MgATPgammA, and MgAMPNP complexes
RT of the Dictyostelium discoidium myosin motor domain.";
RL Biochemistry 36:11619-11628(1997).
RN [8]

```

RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
 RA MEDLINE:98070605; PUBMED:9405148;
 RX Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.,
 RT "X-ray crystal structure and solution fluorescence characterization
 of Mg₂.(3')-O-(N-methylanthraniloyl) nucleotides bound to the
 RL Dictyostelium discoidium myosin motor domain.";
 J. Mol. Biol. 274:394-407(1997).
 CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
 CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
 CC -1- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
 CC INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
 CC LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
 CC (MLC-2).
 CC -1- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
 CC CORTEX.
 CC -1- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEMOROSIN (LMM) AND 1 HEAVY MEMOROSIN (HMM). IT CAN BE FURTHER
 CC SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
 CC THE ACTIN-ACTIVATED ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDPA ATPASE
 CC ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CTS AT THE SH-1
 CC POSITION (688).
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M14628; AAA3227.1; -
 DR PIR: A26655; A26655.
 DR PIR: S00250; S00250.
 DR PDB: 1MMA: 03-DEC-97.
 DR PDB: 1AMD: 17-AUG-96.
 DR PDB: 1IMG: 03-DEC-97.
 DR PDB: 1MMN: 03-DEC-97.
 DR PDB: 1MND: 17-AUG-96.
 DR PDB: 1MNE: 17-AUG-96.
 DR PDB: 1VOM: 23-DEC-96.
 DR PDB: 1LVK: 28-JAN-98.
 DR DictyDB: DD01008; mhca.
 DR InterPro: IPR000048; IQ.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00612; IQ; 1.
 DR Pfam: PF00063; myosin_head; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR ProDom: PD000355; myosin_head; 1.
 DR SMART: SM00015; IQ; 1.
 DR SMART: SM00242; MYSC; 1.
 DR ProSITE: PS50096; IQ; 1.
 KW Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
 KW Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
 FT DOMAIN 1 761
 FT DOMAIN 762 791
 FT NP_BIND 817 2116
 FT NP_BIND 179 186
 FT DOMAIN 638 660
 FT DOMAIN 738 752
 FT MOD_RES 130 130
 FT MOD_RES 678 678
 FT MOD_RES 1823 1823
 FT MOD_RES 1833 1833
 FT MOD_RES 2029 2029
 SQ SEQUENCE 2116 AA; 243871 MW; 2FC370B1E56A1 CRC64;

Query Match 4.1%; Score 162; DB 1; Length 2116;
 Best Local Similarity 19.2%; Pred. No. 1.8;
 Matches 156; Conservative 127; Mismatches 296; Indels 232; Gaps 38;
 2 KRRKVLPLMALSTILVSTGSLVIAQAEVQENRLNLSSESSGGLGYFSDINFGAP 61
 1200 OKKKVELDELDKSAQAEETAKQALDKLKKLEQELSEVQTO-----LSEANNK-- 1249
 62 MYVTSSTGDDLSIPSELENIPESENOYQSAIWSGFIYKKKDEYTFATSDNHTMWD 121
 1250 NVNSDSTNKKHL--ETSFNNIKLELEAEQK-----KQALEKKRIKLESELKHVNEOL 1300
 122 DQEVINKASNSKIRLEKQALQIKYQYORENPTKGLDFLYWPDSONKKE----- 173
 1301 EKKKQESNEKKKVDLEK-EVSELDQIEEYASKAV-----TEAKKKSELEDEIKR 1353
 174 ---VISSDNLQLPKIKOKSSNRKRRSTSAQTPVDRDNDGIPDSLEVEGYTVDKNR 229
 1354 QYADVSSRDKSVEQLTKLQAKNEELRNTA-----EEAQGLDRAERS-----KKKA 1400
 230 TF-LSPWISNIEKKGLTKYKSSPKWSTADPSYDFEYVGRID--KNVSPEARHPLVA 286
 1401 EFDLEBAKALEE---TAKKAKAEAKKAAE--TDYRSTSELDADKAVSEQ----- 1449
 287 AYPVHVDMENITLSKNEDOS-----TONTDETRTISKNTSTSRTHSEVGNAEVHAN 341
 1450 ---YVQIKRL---NEELSELRSVLEBADERCSAIKAKKTAFSALESLELDEIDANN 1500
 342 TSTSRTHSEVGNAEVHANVALDHSLSLAGEETMAETGCLNADTARLANRINYNTGTA 401
 1501 AKAKAKERSK---ELEVRVALEESLE-----DKSG-----TVNVEFIRKDA 1540
 402 PIYVNLPTSTVLGKNQTLATIKAKENQSLIAPNNYPSKNLAPIALMAODDESPPI 461
 1541 ELDIDLARLD-----RETFSRIKSDK-----KN----- 1565
 462 TNNVQFLEKTKQRLDQVYNIAATYVNEGRRVADTGSNMSEVLPQIO-ETTARI 520
 1566 --TRKQFADL-----AKVEAQREVVITD---RLKKLESIDILIDSLQDTETKSR 1613
 521 -TFNCK---DNLVERRRIAANVPSPLETTRKPDMLKEAKIAFGFNPNGMLQYOGNDI 576
 1614 KTEKSKKKLEQTLERRAAEBGSSRAAD-----EETRQVW-----OEY 1652
 577 TFEFNFDOQ-----TSQNIKQALAEINATNITYVLDKIKLNAKMILLRDKRPHYDRN 630
 1653 DELRAQLDSERAAALASEKKIKSLVAE-----VDEKQLEDELLARDKLVAKARA 1703
 631 ---NIAVGADESVEKHAHREYVINSSTEGILLNIDKDIRKILSGYIYEI----- 676
 1704 LEVELEEVYDQLEEEEDSESELSKRRRLTTEVE-DIKK---KYDAEVEQNTKLDEAKKK 1759
 677 ---DREGLEKEVINDRYDMNITSS-----LRQKGFY-----DFKKYDK 713
 1760 LTTDDVDTLKKQLEDEKKNLSEERAKKRLESSENEQFLAKDAEVKRRRAEKDRKYEKD 1819
 714 LPLYSINPNYKVVYAVTKENIINPSENGD 744
 1820 L-----KDKRYKLNDAARAKTKQTEIGAANKLED 1846
 RESULT 9
 BBP1_PLAVB STANDARD; PRT: 2869 AA.
 ID BBP1_PLAVB 000798;
 AC 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
 GN BBP1.
 OS Plasmodium vivax (strain Belen).

Db 655 TLSDNNDLRTKLLKLEESNLSLIKQEDVDSLEKNIQTLKEDLRKSEALFSLKAKN 714
Qy 185 LKOKSSNRKRRSTGAGTPVDPDNDGIPDSLEVGYTYDVYKNRTPSPWISNHEKKG 244
Db 715 LREVIDNLKGRHETL-----EAQRNDLHSSLS-----DAKNTAIISSLELT----- 755
Qy 245 LTKYKSSPEKMSSTASDPYSDPEKVTGRI-----DKNVSPEARHPLVAAPYV----- 291
Db 756 ----KSS-----EDVKRLTANVETTLQDQSKAMQSPSTSLVNSYQSISNLVHELR 800
Qy 222 --HYMEMIILSKNEDOSTQUNDSTFRITKTSRHTSRYHNAEYHANTSRHT 349
Db 801 DDHVMQSONNTLSESESKLTKDCENL-----TQONMTLIDVQKRLMHVQESKVSLE 855
Qy 350 SEVHNAEYHVAHAIDHSLSLA-----GERTMAETMGLN-----TADTANANIRIVN--- 397
Db 856 KEVNKSLDLKLNLSLNVASISDNDQILTOALSLKSNVDSLEQSAQNSGLKLEAK 915
Qy 388 -----TGAPYIVLPTTSLVGLKNOPLATIKAKENOLSLAPNNY 440
Db 916 QLLHTNEBELHRLDKLTKLIEE-----SKSDLGKLT-----AROEISMLKEEN--- 964
Qy 441 PSKNLAPIALNADDESSPTITMNVNQLFLEKTKQLRDLTDQVYGNATYNE----- 494
Db 965 MSQSAITSVSKLDET-----LSKSKLEADIEHLNKNVSEVEVERNALLA 1011
Qy 495 -NGRVVD--TGSNMSEVLPOIOTTAIIIFNGKDLVERRIAVNSPDLFTTKPDM 550
Db 1012 SNERLMDLKNNGENIASIQETIEKKRAE-----NDLQ--SKLSV 1050
Qy 551 TLKELKLAFGNEPNGMLQOGKITEEDNR---PQOQTSQNIKNOLAELINATNYTV 606
Db 1051 VSSEYENLLISSQTNKSLKTKNOLKYEKVNKLDKEDQDRN--ELEEL--TSKYGK 1106
Qy 607 LDKIKLAKMNLIL--BDRPH-----YDRNNIANGAESVYK 642
Db 1107 LGEENAOIKDELALRKSKKHODLCANFVDDLKESKSALEOLINEKVELIVSLEOS--- 1163
Qy 643 EAHREVINSSTEGT-----LLNIDKIRKILSGYIVEIEDEGLEKEYI-----NDRY 689
Db 1164 -----NSNNEALVEERSDLANRLSDMKKSL-----DSDNVSIYRSLAVANDEL 1209
Qy 680 DML-----NISLQODGTFID-----FKYNDKILPLYSINPNKVVNYAVTK 732
Db 1210 DTLKKDKSLSTQYSEVQDRDDLDSLKGCESEPNKYAVSLRELCTSEIDVPSLID 1269
Qy 733 ENTINPSENGDTPNGIKILIFSKKGY 761
Db 1270 DNFVNAGNFSELS-----RLTVLSLENY 1293

RESULT 11
USOL_YEAST STANDARD: PRT: 1790 AA.
AC P25386:
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTRACELLULAR PROTEIN TRANSPORT PROTEIN USOL.
GN USOL OR INT1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
protein transport in Saccharomyces cerevisiae.";

RL J. Cell Biol. 113:245-260(1991).
RN (2)
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -!- DOMAIN: THE RODLIKE TAIL. SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -!- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: X54378; CAA38253.1; -;
DR EMBL: L03188; AAB00143.1; -;
DR EMBL: U53668; AAB66559.1; -;
DR PIR: A38455; A38455.
DR HSSP: P80220; IDIP.
DR SCD: S0002216; USOL.
DR InterPro: IPR002017; Spectrin.
KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724
FT DOMAIN 725 1790
FT DOMAIN 465 1487
FT DOMAIN 991 1790
FT DOMAIN 1172 1786
FT CONFLICT 847 847
FT CONFLICT 924 924
FT CONFLICT 1253 1253
FT CONFLICT 1319 1319
FT CONFLICT 1461 1461
FT CONFLICT 1581 1581
FT CONFLICT 1600 1600
FT CONFLICT 1661 1661
FT CONFLICT 1772 1772
SQ SEQUENCE 1790 AA; 206424 MW; 6CEB216E9FD4818 CRC64;

Query Match 4.1%; Score 160; DB 1; Length 1790;
Best Local Similarity 18.2%; Pred. No. 1.8;
Matches 142; Conservative 145; Mismatches 288; Indels 206; Gaps 32;
Qy 23 NLEVIOAEVKQENRLNSESQGLGYFSDLNFOAPMYVTSITGD-----LSI 74
Db 1023 NIEQKKTISDLEQKEKELISKDSKDEYSQISLKEKLETTATANDENVNKISL 1082
Qy 75 PSSELE-----NIPSENYQFQSAIWSGFYKKSDEY-----TPATSADNHVTM 118
Db 1083 TRELEAEIAAYKNKLNLEKLETSERALKVEKNEHEKLEKIQLEKEATEETKQOLNS 1142
Qy 119 WVDQEVINKASNSMKIKLEGRLYQIKI-----QY-----QRENPTEGL 159
Db 1143 LRANLELEKEHEDIAOLK--YEEQIANKERYNEISQLANDEITSQOENESIKK 1199
Qy 160 DFKIYW-----TSQKKKEVYSSDNLQPLKOKSSNRKRRSTGAGTPVDPDN 209
Db 1200 NDELEGEVKAKKSTSEBSNLSKSEIDALNLQIKRLKKN-----ETNE 1243


```
Db 1547 LDE-----RECKCLLVNKKQCGDKC---VENPNTCN-----EN-----NGG 1580
OY 744 -----DSTNGIKKI 753
Db 1581 CDADAKCTEEDSGSNG-KKI 1599

RESULT 14
MDS3_YEAST STANDARD: PRT: 1487 AA.
ID MDS3_YEAST STANDARD: PRT: 1487 AA.
AC P53094;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MDS3 PROTEIN (MCK1 DOSAGE SUPPRESSOR 3).
GN MDS3 OR YGL197M OR G1307.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;

RN SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97197971; PubMed=9046087;
RA Coglielva M., Klima R., Bertani I., Delneri D., Zaccaria P.,
RA Buschi C.V.;
RT "Sequencing of a 40.5 kb fragment located on the left arm of
RT chromosome VII from Saccharomyces cerevisiae."
RL Yeast 13:55-64(1997).
RN [2]
RP CHARACTERIZATION.
RA Li M.B., Neigeborn L.;
RL Unpublished observations (XXX-1997).
CC -I- FUNCTION: NOT KNOWN; NEGATIVE REGULATOR OF EARLY MEIOTIC GENE
CC EXPRESSION.
CC -I- SIMILARITY: TO YEAST YER132C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X91837; CAA62947.1;
DR EMBL; 272719; CAA96909.1;
DR SGD; S0003165; MDS3.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF01344; Kelch; 2.
KW Melosis.
SQ SEQUENCE 1487 AA; 167073 MW; 768AEFBAB796E447 CRC64;

Query Match 4.1%; Score 158.5; DB 1; Length 1487;
Best Local Similarity 19.1%; Pred. No. 1.7; indels 255; Gaps 42;
Matches 170; Conservative 145; Mismatches 318;

OY 14 STILVSTGNLEVIOAEVQENRLNESSSQGLGYFSPDNPQAPMVTVSS----- 67
Db 593 STILKSPHSSRNSSKAVKQEGRL-----SSGSLDNVF-EKNF--PIFARTSVSEADN 643
OY 68 TTGDLSIPSELENIPESENYQPSAIVSGFIKVKSDSEYTFATSDNHVTMNV----- 120
Db 644 TQPOVANADAKAPNTPS-----TSDSESSSSSDLYSTPHYQRNDDE 686
OY 121 DDOEVYN--KASNSKIRLEKGRLYO--IKIQYQRENPEPEKGLDFKLYWTD--ONKKE 173
Db 687 DDEDPVSPKPVSKNSI-----YPIRKTESSSTTSSNGMIFVPEFEKAAYTSNTE 739
OY 174 VISSDNLOLPELKOKSSN-----SRKKRSTAGPTVPDRNDGTP-- 213
```

```
Db 740 ALLESNLSIQLESRRRSLMSIPSGELLRSSISEAEHORASHPLTSSPLFEDSGTPCGK 799
OY 214 DSLVEGTYVDVKNKRTFLSP-----WISNHEKKLTLYKSSPEKMWASDP 261
Db 800 QLOOLOOQHT--IONPHNHLSPPRFRSARSRSISYVSSSDRRG-----NSISSRT-SDS 851
OY 262 YSDFEKVYGRIDKNVSPKARHP-----
Db 852 FGT-PPVLGVNLVPLPPQTRPEEPPPPCPAMSTGSTRSRNTLTDMISNKAAPSSRR 910
OY 284 -----LVAAPYIVHMENTIISKNEQSTQNTDSETRTISKNTSTSRTH 328
Db 911 SSHIGRRSSPETENAFSAPPRASLDGQMGKSLKSGSTQYQPRKNSPKANETIQTP 970
OY 329 TSEVHGNAEYHANTSTSRHT-SEVHGN--AEYHAVAIDHSL-----AGEPRTAETMGL 381
Db 971 TS--SNNEWSQSVTSNDSFDSLOSNFALDEPLLTSPSLVMPWPTSTVRAFAEFYPT 1027
OY 382 NTADTARLNINIRYVMTGTAPIYVNLPTTSLVGLKNQTLATIKAKENQSLQILAPNNYYP 441
Db 1028 GOVNSKWLAPVALDLVMAKITEIPLKYLIL--EVLSTLAKKESLSICTSLMET 1084
OY 442 SKNLAPIALNAODFSSPTTM--NNQPLELEKT-----KQLRLTDQV 484
Db 1085 FRRTKLNSYKGEDEKNTVLTSDNDYQELLKLVSLNIDNGYVDPDLRKQSRQSSST 1144
OY 485 YGNIAIYFNENGRVYDGTGNSWSEVLPQIOETTARIIFNG-----KDLNL 529
Db 1145 QESGSGSANGF---KTAGAGSLET---SSTWVPYFAGGPRDSHNSVSGISGFPNSMNI 1196
OY 530 VERRIAAVNPDDLETTKPPMTLKEALKIAFG--FNEPKNLOYQSKDITDEFNDQ 586
Db 1197 QGSRRTSGFS-----PVKMKSSLSKEIDPKTFYE---EYEPKEKSPDDNDQ 1244
OY 587 TSONIKN-QLAELN-----ATNIYT-----VLDRIKLAKKNNILIRDKRFH 626
Db 1245 YNIGSFRLHLFDNNYSGISSSTNSISSDLEKEBEQOJDLLEIREDSAELIDARF- 1303
OY 627 YDRNNIAVGADESIVYKAEHREVINSSTEGILLINDIKRILSGYIEIDTEGLKEVIN 686
Db 1304 --RNK-----EDDVTYKQ-----ISNDKRNLYLPHEKNNLKAKKG--KETRDVREEEEDF 1350
OY 687 DRDMLNSSLRODGKTFIDFKKYNDKL-PLXISN--PNKKVNVYANT 731
Db 1351 FGLGMLSLNKKIKREAK-HVD--KYDSDVPLFSSAPFQSPIRAYGST 1395

RESULT 15
FAT1_SCHPO STANDARD: PRT: 1385 AA.
ID FAT1_SCHPO STANDARD: PRT: 1385 AA.
AC O13735; Q90UTJ0;
DT 15-JUL-1998 (Rel. 36, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ACTIN INTERACTING PROTEIN 3 HOMOLOG.
GN FAT1 OR SPAC15A10.15 OR SPAC15E1.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RN SEQUENCE OF 1-1033 FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 1023-1385 FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Lyne M.H.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RN IDENTIFICATION, AND GENE NAME.
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:51:05 ; Search time 72.22 Seconds

(without alignments)
805.834 Million cell updates/sec

Title: US-09-747-521-4

Perfect score: 3913

Sequence: 1 MKRRKVLIPALSTILVSS.....TSTNGIKKILFSSKKGYEIG 764

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database :

1: PIR68:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3800	97.1	764	2	I39934 protective antigen
2	883.5	22.6	875	2	I40862 iota toxin compo
3	235.5	6.0	192	2	I39933 cryptic protein -
4	235.5	6.0	204	2	G59104 hypothetical prote
5	205.5	5.3	4688	2	F82885 hypothetical prote
6	193.5	4.9	4919	2	T31105 hypothetical prote
7	193	4.9	1072	2	A86827 hypothetical prote
8	190	4.9	1302	1	JC6009 surtface-located me
9	186.5	4.8	2401	2	T28676 rhopary protein -
10	186	4.8	2178	2	S55805 alpha-toxin - Clos
11	183	4.7	1465	2	S31262 TYB protein - yeas
12	183	4.7	1803	2	S56894 TYB protein - yeas
13	182.5	4.7	1193	2	S68218 botulinum neurotox
14	182	4.7	1658	2	S55101 hypothetical prote
15	180	4.6	1051	2	T18351 ImpI protein - Myc
16	180	4.6	2269	2	T28677 rhopary protein -
17	179.5	4.6	1365	2	T30822 ImpI protein - Myc
18	177.5	4.5	1467	2	PC1253 TYB protein - yeas
19	177	4.5	1315	2	T28679 fibrinogen-binding
20	177	4.5	1837	2	T41023 probable nuclear p
21	177	4.5	3724	2	T18427 hypothetical prote
22	176	4.5	1620	2	S61535 nucleotide-binding
23	175.5	4.5	1516	2	E71619 RAD2 endonuclease
24	175	4.5	1802	2	S52611 TYB protein - yeas
25	175	4.5	1939	2	T18372 repeat organellar
26	174	4.4	2529	2	B64635 toxin-like outer m
27	173.5	4.4	1230	2	S56850 SMCI protein homol
28	173.5	4.4	1271	2	D64237 hypothetical prote
29	173	4.4	1230	2	E64664 outer membrane pro

30	171.5	4.4	821	2	S67087 hypothetical prote
31	168.5	4.3	1115	2	T41342 probable coiled-co
32	168.5	4.3	2526	2	T20531 hypothetical prote
33	168	4.3	2523	2	T18477 hypothetical prote
34	168	4.3	4152	2	T31102 filamentous hemagg
35	167.5	4.3	786	2	T18469 ORF MSV156 hypothe
36	167	4.3	1127	2	T28317 surface membrane p
37	166.5	4.3	624	2	PC6003 myosin heavy chain
38	166.5	4.3	1928	2	S46773 hypothetical prote
39	165.5	4.2	1553	2	T18502 fibrinogen-binding
40	165	4.2	1166	2	T28680 hypothetical prote
41	165	4.2	5005	2	F82884 p101 protein precu
42	163.5	4.2	888	2	A38539 protein with DnaJ
43	163.5	4.2	1014	2	H71602 toxin-like outer m
44	163.5	4.2	2399	2	H71879 cell surface antic
45	162.5	4.2	2340	2	B71704

ALIGNMENTS

```
RESULT 1
I39934
protective antigen precursor - Bacillus anthracis plasmid
C:Species: Bacillus anthracis
C:Date: 19-Jul-1996 #sequence,revision 19-Jul-1996 #text,change 01-Dec-2000
C:Accession: I39934; S69160; F59104
R:Weikos, S.L.; Lowe, J.R.; Eden-McCutchan, F.; Vodkin, M.; Leppla, S.H.; Schmidt, J.
Gene 69, 287-300, 1988
A:Title: Sequence and analysis of the DNA encoding protective antigen of Bacillus ant
A:Reference number: I39933; MUID:89172073
A:Accession: I39934
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-764 <RES>
A:Cross-references: GB:M22589; NID:g143280; PIDN:AAA22637.1; PID:g143282
R:Fieldman, T.C.; Gordon, V.M.; Leppla, S.H.; Kimpel, K.R.; Birch, N.P.; Loth, Y.P.
Arch. Biochem. Biophys. 316, 5-13, 1995
A:Title: In vitro processing of anthrax toxin protective antigen by recombinant PC1 (
A:Reference number: S69160; MUID:95142670
A:Accession: S69160
A:Molecule type: protein
A:Residues: 197-202 <FR>
R:Okinka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koeh
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pXOI, the large Bacillus anthracis plasmid harb
A:Reference number: A59091; MUID:99445483
A:Accession: F59104
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-313, 'O', 315-764 <OKI>
A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32414.1; PID:g4894326
A:Experimental source: strain Sterne
A>Note: similar to anthrax toxin moiety, protective antigen, paga formerly pag, plasm
C:Genetics:
A:Gene: pXOI-110
A:Genome: Plasmid
C:Function:
A:Description: three component exotoxin; protective antigen binds to receptors on the
y active components edema factor or lethal factor; the complex is internalized by rec
C:Keywords: exotoxin
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-196/Domain: propeptide #status predicted <PRO>
F:197-202/Product: protective antigen #status experimental <MAT>
```

Query Match 97.1% Score 3800; DB 2; Length 764;
Best Local Similarity 97.6% Pred. No. 8.1e-186;
Matches 746; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 MKRRKVLIPALSTILVSSGTNGEIVIOAEVKEORLLNSESSESSQGLGYFFDLNFOA 60
DB 1 MKRRKVLIPALSTILVSSGTNGEIVIOAEVKEORLLNSESSESSQGLGYFFDLNFOA 60

QY 61 PAVVYSSTTGGDISIPSSSELENIPSENOYFQSAIWSGFIYVKSDEYTPATSAHNTMMV 120
|||||
Db 61 PAVVYSSTTGGDISIPSSSELENIPSENOYFQSAIWSGFIYVKSDEYTPATSAHNTMMV 120
QY 121 DDOEVYINKASNSNKRIRLEKGRLYQIKIYOORENPTEKGLDFKLYWTDSONKKEVISSDNL 180
|||||
Db 121 DDOEVYINKASNSNKRIRLEKGRLYQIKIYOORENPTEKGLDFKLYWTDSONKKEVISSDNL 180
QY 181 QLPETLKQSSNRKRKRSTASGPTVPDRONDGIPDLSLEVGGYVYDVANKKRTFLSPMISNTH 240
|||||
Db 181 QLPETLKQSSNRKRKRSTASGPTVPDRONDGIPDLSLEVGGYVYDVANKKRTFLSPMISNTH 240
QY 241 EKKGLTKYKSSPEKSTASDPYSDFEKTVGRIDKANSPEARHPVLAAYPIVHDMENIIL 300
|||||
Db 241 EKKGLTKYKSSPEKSTASDPYSDFEKTVGRIDKANSPEARHPVLAAYPIVHDMENIIL 300
QY 301 SKNEDQSTQNTDSEFRTISKNTSTSRTHSEVHGNAEVAHANTSTSRTHSEVHGNAEVAHA 360
|||||
Db 301 SKNEDQSTQNTDSEFRTISKNTSTSRTHSEVHGNAEVAHANTSTSRTHSEVHGNAEVAHA 360
QY 361 VAIDHSISLAGRTAAETMGLTADTARLANANIRYVNTGTAPRYNVLPTTSLVLGKNQTL 420
|||||
Db 361 VAIDHSISLAGRTAAETMGLTADTARLANANIRYVNTGTAPRYNVLPTTSLVLGKNQTL 420
QY 421 ATIKAKENOLSGILAPNNYPPSKNLAPILANODDESSPITNNYNOFLEKTKQRLD 480
|||||
Db 421 ATIKAKENOLSGILAPNNYPPSKNLAPILANODDESSPITNNYNOFLEKTKQRLD 480
QY 481 TDQVYGNATYATYFENGVRNVDTGSMWSEVLPOIETTARIIFNGKDLNVERRIAANVPS 540
|||||
Db 481 TDQVYGNATYATYFENGVRNVDTGSMWSEVLPOIETTARIIFNGKDLNVERRIAANVPS 540
QY 541 DPLETTKPDMTLKEALKIAPGFENEPNGNLQYOGKOTIEEDFNDDOOTSNIKQOLAEINA 600
|||||
Db 541 DPLETTKPDMTLKEALKIAPGFENEPNGNLQYOGKOTIEEDFNDDOOTSNIKQOLAEINA 600
QY 601 TNYTVLDKIKILNAKNMILIRDKRFHYDRNNIYAVGADSEVYEAHREVIINSSTEGLLNT 660
|||||
Db 601 TNYTVLDKIKILNAKNMILIRDKRFHYDRNNIYAVGADSEVYEAHREVIINSSTEGLLNT 660
QY 661 DKDIRKIISGTYVEIEDTGTGLKEVINDRYDMLNISLRDQKTFDFPKKYNKDLPLYISN 720
|||||
Db 661 DKDIRKIISGTYVEIEDTGTGLKEVINDRYDMLNISLRDQKTFDFPKKYNKDLPLYISN 720
QY 721 PNYKVNVAVTAKENTIIINPSENGDSTNGCIKKILIPSKKGYETG 764
|||||
Db 721 PNYKVNVAVTAKENTIIINPSENGDSTNGCIKKILIPSKKGYETG 764

RESULT 2
140862
lota toxin component Ib - Clostridium perfringens
C:Species: Clostridium perfringens
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Oct-1999
C:Accession: 140862; 542774
R:Perelle, S.; Gilbert, M.; Boquet, P.; Popoff, M. R.
Infect. Immun. 61, 5147-5156, 1993
A:Title: Characterization of Clostridium perfringens lota-toxin genes and expression in
A:Reference number: 140861; MUID:94041637
A:Accession: 140862
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-875 <RES>
A:Cross-references: EMBL:X73562; NID:g929031; PION:CAA51960.1; PID:g414655

Query Match 22.6%; Score 883.5; DB 2; Length 875;
Best Local Similarity 31.5%; Pred. No. 1.8e-37;
Matches 262; Conservative 135; Mismatches 279; Indels 155; Gaps 35;

QY 15 TLVVSSTGNLEY-----TAAEVKQENRLNISESSSQGLGTYFSDLNFOAPMVVYSSTT 69
|||||

Query Match 6.0%, Score 235.5; DB 2; Length 192;

C.Accession: T31105
R.Ward, C.K.; Lumbley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.
J. Bacteriol. 180, 6013-6022, 1998
A.Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
A.Reference number: 220984; MUID:99030326
A.Accession: T31105
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-4919 <MAR>
A.Cross-References: EMBL:AF057696; NID:g3929021; PID:g3929023; PIDN:AAC79761.1
C.Genetics:
A.Gene: lspa2

Query Match 4.9%; Score 193.5; DB 2; Length 4919;
Best Local Similarity 21.1%; Pred. No. 0.27;
Matches 180; Conservative 131; Mismatches 324; Indels 219; Gaps 42;

1 MKRKVLIPMLSTLSTVSTGTLVIAEYKQENRLNLESESSSGGLGYFSDLNFGQA 60
245 VORGVYIKKDVAT---NGLSFDVAAKNIDQGVKSIEGSKPAKL---ANVTFAA 296
61 PMVYASSTTGDLSPSELENIPEENQFQSAIWSG-----FIKYKSDSEYF 108
297 GMLTYDVNTRDVNRNTPKPKITDNRKDNIAI-SCESAGSMGRIKIVTDKG----- 350
109 ATSAQNH--VTMMVDQEVINKASNSNKIRLEKGRLYQIKIYQRENPTEKGLDFKLYWT 166
351 --AGVHHQVIAEDDINILITDDGNS-----RLNKYADYVR---VGGKIDEL--- 393
167 DQONKEVSSNLOLPELKOKSSNRKRSAGPTVPDRNDGIPD-SLEVEGYTVYV 225
394 -----ANNQGHADQDLINATGTVKLVNDGSSVSNLNNLISALNLTLENAVSA 443
226 KKKRTFLSPWISNIHEKKGLTKYKSSPEKWSASDPYSD-----FEKYTGRIKN 275
444 NN-----LSFRVTNDRKLNLSKVS-----ARAADQSGNMLDKRASVLAHLTLTINS 493
276 VSEPARHPLVAAPYIVHDMENIILSKNEDOSTONTSETRTISKNTS--TSRTHTSEV 332
494 VSLNNOKSLA-----NNLKIKYVRDLNMLNSELANNLTLTSSNITLKNKSKPT 544
333 HONAEVHANTSTSRTHTSEVHNAEVAHAVIDHSLAGEKRTWAETMGNTADTARLAN 392
545 AGNMTLNTANNVTLNDSFLAANNLTLANTKVNTLNDASKLS-ANKLDLNTVDNVTLN 603
393 IRVYVGTAPIYVNLPTTSLVGLKNOTLATIRAKENQLSQILAPNNVYPSKMLAPIALVA 452
604 -STLSAGELTFKKV-----KNVTL-----NNSE-LAANN-----LSLNA 636
453 QDDFSSTPTMNYNOFLEKLT--KQRL--DTPQVYGNATVYFENGVRVDTGSMWS 507
637 ---SHVYTLNKKSKLSQKADIAVNLTLNDTTE---LTKAKMLDINSTITNNGTIA 687
508 EYLPQIOETTARIIFNGDLNVERRI-AAVNPSPLETTKRDMLKELKATAPFENEN 566
688 GIFANI--TTEKLNKKKALILAEONLFTYNGSH--YENKQDIYSKDAVATYTESKNS-- 741
567 GNLQYQKGDITEEDFNPDQOSTONIKQIQAELANATNI-----TVLDIKILANK--- 615
742 -----DFTSGSKLVNAQNDL-KVYVNNFTISSQDDITLLIGNVTLNASSGFTT 787
616 -----MNTIIRKRPYHRRNNTIAGVAD--ESVVK-----EAHREYI 649
788 NSGNTLTVTKTLVDGIQNTKGNLTVEEDLHKSKTKITNDGKLISIKNLINSEADPT 847
650 NS-----STREGLLNIDKDIRKILISGYIEIDTEBKLKLVINDRYM---LNI 695
848 NNGTLLGIEALKIATKGNFTNKEKAI--LASNSLDISVAEKKKFTNNGTIESGKNLNT 905
696 S-----LRQDGTFIDFKKYNDKLPYISN-----PNYKYNVAVKEMTIINSPSGDS 746
906 MTGAFNLVNDMNTIRSFVGLNITSTGVNSNNGTLISNERLN---ITSANFTNES-NGTVM 961

RESULT 7
A:Status: preliminary
A.Molecule type: DNA
A.Residues: 1-1072 <STO>
A.Cross-References: GB:AE005176; NID:g12724625; PIDN:AAK05715.1; GSPDB:GN00146
A.Experimental source: strain IL1403
C.Genetics:
A.Gene: yqfg

Query Match 4.9%; Score 193; DB 2; Length 1072;
Best Local Similarity 18.4%; Pred. No. 0.03;
Matches 152; Conservative 155; Mismatches 341; Indels 176; Gaps 30;

7 LIPMLALSTILVSTGTLVIAEYKQENRLNLESESSSGGLGYFSDLNFGQAPMVYTS 66
319 LYPISSEASVTNLTNLSISLDSITS-----SQTENSGSG-----A 355
67 SPTGDLSPSELENIPEENQFQSAIWSGFIKYKSDSEYTTATSAQNHVTMMVDQEVY 126
356 SSTAEISTDSENSNSLSSNQ-----INSNSNEKDSNOSLSGSSWSSESEHSNSNI 410
127 NKASNSNKIR-----LEKGRLYQIKIYQRENPTEKGLDFKL-----YMTDSQNKKEVSS 177
411 NCTNNSSETITNLLPSPNPTESNSVDQSTSEASTNSNSISLSPNISSTDSSEATNS 470
178 DNLQLEPELKOKSSNRKRSAGPTVPDRNDGI---PDSLEVEGYTVYVKKRTEFLS 233
471 DFNVAEVAVNNSLASVNNSSSVLSTSTADNLGINQSGDNLTGKD--SEISTGAFLS 528
234 PWISNIHEKKGLTKYKSS-----SPEKMS-----TASDPYDFEYKVRIGRIDKNVSPKAR 281
529 ---SNQTSSEASTNSNSISLSPNISSTSVLESTTSSNFSNVAEVAVNNSLASVNNSSS 585
282 HPLVAAPYIVHYDM-----ENIILSKNEDQ-----STONTSEYRTISKNTSTSRTHTS 330
586 SVLSTSTADNLEINOFPGSDNLTKQSEISTSGAFLLSNQSTSEASNSMSINSPLSL 645
331 EYHGAEVHANTSTSRTHTSEVHNAEVAHAVIDHSLAGEKRTWAETMGNTADTARLN 390
646 SLTNSSESATNOSNS-SEATKYDNNSTHSSNILLNS-----GSDSDSDSDSDSDSSNLT- 699
391 ANIRVYVGTAPIYVNLPTTSLVGLKNOTLATIRAKENQLSQ---ILAPNNVYPSKMLA 446
700 -----SSSPNLETNQTISKPSFVNNNISPPKPVSSNSVQENSTIDH 741
447 PIALNAODFSSTPTMNYNOFLEKLTQRLDTPQVYGNATVYFENGVRVDTG 503
742 EMSTPKSISISPIITSSSQOKESQSN---LLNTTEGINNPTITNNSSEKASISILT- 797
504 SNMSEVLPQIOETTARIIF-----NGKDL-----NLVERRIAAVNPSPLETTKPKDM 550
798 -SYSNNNSSEPTGCLYISNEAQRNGSEISHLPPSSNENNVSSIOSQAILLESKSSST 856
551 TLK-EALKIARFENP-NGNLQYQKGDITEEDFNPDQOSTONIKQIQAELANATNIYVLD 608
857 NKRSSLSIINSTSHPNEDQSNSSD-----EVKSNNVESILQOLNLSISKNTHMN 908

[illegible]

```

OY      VITGAPPIYNVLPTSLVGNKOPLATIKA---KENOLSQLAPNNYPSPKNLAPIALNA 452
Db      810 ANTLAKL-----KDNITQQAQTELEKEVKQANAVASNNTASQSAKSILDA 859
OY      453 Q-DDFSPTITMWNQFIELEKTOKLRIDTDOVYGNIATVPNFENGRRVADTGSMSEVLP 511
Db      860 KVTETTKLETFENKKDKVFKFELEGTRKDIDEF-----INTNKTNPDYST 904
OY      512 QIQEFTARIIFNGKLNIVERRIAIVNPSPDLETTPKDMTLKEALKIAFGNEPENGNOY 571
Db      905 LISELTSK-----RDSKNSITNSSNSDLETATTELKOALAFA-----942
OY      572 QGKDIITEFDNFDDQTSNIKQLAELNATNYIYLVIKIKINAMNIIIRDREHYDRNN 631
Db      943 -NTDKDAQD-NLARSTRQLNKSISAN-----FLAALT-----DKDN 979
OY      632 IAVGAD---ESVYKAHAEVINNSTEGLLINDIRKILESGIYEIEDTGCKEVINDR 688
Db      980 TIQAQTELEKEVKQANAVASNNTASM-----OSAKSSLDAAKTEI--TKLETFPNKD 1032
OY      689 -YDMAINISLRDGCKTFEDFKFYNDKLPLYISPNPYKANVAVT---KENTIIINSPEG 743
Db      1033 DYKFEELEGTRKDIDEFTINTRK-----TNPYSLISELSISKRSKSNSTNSSKKS 1083
OY      744 DTST 747
Db      1084 DIET 1087

RESULT   9
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C.Species: Plasmodium yoelii
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C.Accession: T28676; A:45521
R.Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A.Title: Comparison of two members of a multigene family coding for high-molecular ma
A.Reference number: 220507; MUID:97077455
A.Accession: T28676
A.Status: preliminary
A.Molecule type: DNA
A.Translation: translated from GR/EMBL/DDBJ
A.Residues: 1*2401 <SIN>
A.Cross-references: EMBL:U36927; NID:g10411784; PID:g10411785; PIDD:AAB41263.1
R.Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A.Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple
A.Reference number: A45521; MUID:91101660
A.Accession: A45521
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 2260-2401 <KEE>
A.Cross-references: GB:M34281

Query Match          4.8%; Score 186.5; DB 2; Length 2401;
Best Local Similarity 21.7%; Pred. No. 0.21;
Matches 182; Conservative 123; Mismatches 313; Indels 219; Gaps 44;

OY      30 EVKQENRLINEESSOGILGYEFSDLN-----EQAPWVTVSTTGDLISI 74
Db      402 EVAKENVQLNVYKSNILEKKHYNDQIMINDIKEKAKONDOFREHKHTIPREMRYOK 461
OY      75 PSSLENIPISENQYQSAI--WSGFKY--KSDYTFATSA DNHYVTMMVDQEV---I 126
Db      462 PSIEIKIMKDE---FLSKVNKYNDPDFKYYKRVESEHNKFTLTINKITKTESDEEIKYE 518
OY      127 NKASNSNRILKEGLYOKIOYORENPFTEGIDPEKLVLTWPSQKKKEVISSDNLQLEPK 186
Db      519 NMFNDSKSLINETKK--SIEEYONINTLKRVDD---YIKVCLNTNELITCNKKQTTLK 573
OY      187 OKSSSRK-KRSTSGAGTVPDRDNDGIPD-----SLEVGYGVADVKKNRFTLSP 234

```

```
Db 574 DKLNONIKITIKETNSIDKIYTKDFENIILDTKTELETCTGLSNHNSNKKELTYEFD 633
Qy 225 WISNIHEKKGLTKYKSSPEKMWSTADSPYSDFEKVYGRIDKNVSPARHPLVAAYPIVHD 294
Db 634 LKANIGKKNENMLYKOFNEK-----EKAVEDIKKKNDVINKIVSN-----IETT 677
Qy 295 MENILSKNEDOSTONTDEFTTISKNTSTSTHTSEVGNAEVNAANTSTSTHTSEVHG 354
Db 678 IYTSIYNINED-----TENE-----IGKSIELLNTKYLE-----KXVAVN-----TJNEIKE 720
Qy 355 NAEVAVAIIDHSLSLAGEFTMAETGMLTADPARLANIRYVNTGAPRYNLTPTSTLV 414
Db 721 KKK-----DYFODEPK-----EKNIKYDEN-----KKNIIDITLNOKI 755
Qy 415 GKN-OTLATIKA-KENQUSOI-----LAPN-----NYP-----SKNLAPIA----- 449
Db 756 DKSIELTEIKKSNENHIDEIKQIDKLKKVPKNKMFNEDPKIEKKIENIVEKIDKKKN 815
Qy 450 LNAODDFSTPTIMYNOFLEKTKQLRLDPOYGNATYATNFENGRRVUTGSGMSSEV 509
Db 816 IYKEIDKLNEISKIENKTSLEKLKNILSYGKSLGNLFLOQIDEEKKAEHTTKAMEA 875
Qy 510 ----LPOIETTARIIFENGKDLN-----VERRIAVNPSPD-----LETT-----KP 548
Db 876 YIIDLDNIKKKSOEI-----EKEMINMDIKMDIHKEMKALINISHDOYKIYHTSKNHEKI 932
Qy 549 DMTLKEALKIAFGNEP-----NGNILOYGKDITEFPNEDQOSTONIKQOLAEALNATNIYT 605
Db 933 SDIRKNSLIIDFSEESYIINDIKKELEKNVLE-----SQNNPTDIINOYLSKIE--NIYN 985
Qy 606 VLDKIKLAKKNILIRDRF--HYDRNNTAVGADESVAEAREVYINSTBEILLNIDKD 663
Db 986 IL--KLNKIKKIKDKVXEYDEIEKNN-----KIKAEISNS----- 1020
Qy 664 IRKILSGYVEIETDEGLKE-----VINDRY-----DMINISLRDQGTIFDKFYNDK 713
Db 1021 -EKI-----ITQKENSLSLKEQSKIKSTIDMYSECIKNIINL-----KTYIVNEKN-- 1069
Qy 714 LPLYSIN-PNRYVNV-----YAVTKENTIIINSENGDSTFN--GIKKILIEFSK 759
Db 1070 INTYERNAEEYONVSNFNENIEMADTKSOYIILIKKNGTNNNDYINIKELKEHKKK 1126

RESULT 10
553805
alpha-toxin - Clostridium novyi (ATCC 19402)
C:Species: Clostridium novyi
A:Variety: ATCC 19402
C:Date: 28-Oct-1996 #sequence revision 08-Nov-1996 #text change 05-Nov-1999
R:Holmann, F.; Herrmann, A.; Habermann, E.; von Eichel-Streiber, C.
Mol. Gen. Genet. 247, 670-679, 1995
A:Title: Sequencing and analysis of the gene encoding the alpha-toxin of Clostridium novyi
A:Reference number: 140834; MUID:95342160
A:Accession: 555805
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-2178 <HOF>
A:Cross-references: EMBL:Z48636; NID:g728537; PIDN:CA488565.1; PID:g755724
A:Accession: S711294
A:Molecule type: protein
A:Residues: 1-15 <HOF>
R:Holmann, F.
submitted to the EMBL Data Library, March 1995
A:Reference number: S71158
A:Accession: S71158
A:Molecule type: DNA
A:Residues: 1-1179, 'LKV', 1183, 'LVTHIGE', 1191-2178 <HOS>
A:Cross-references: EMBL:Z48636; NID:g728537; PIDN:CA488565.1; PID:g755724
R:Holmann, F.; Habermann, E.; von Eichel-Streiber, C.
submitted to the EMBL Data Library, July 1993
A:Description: Sequence analysis of Clostridium novyi alpha-toxin: a member of the fami
```

```
A:Accession: S44273
A:Molecule type: DNA
A:Residues: 1-243;1204-2178 <HOF>
A:Cross-references: EMBL:Z23281
C:Genetics:
A:Gene: tcn-alpha
C:Superfamily: cpl repeat homology
C:Keywords: virulence factor
F:1880-1899/Domain: cpl repeat homology <CP12>

Query Match 4.8%; Score 186; Db 2; Length 2178;
Best Local Similarity 20.8%; Pred. No. 0.19;
Matches 165; Conservative 116; Mismatches 264; Indels 250; Gaps 39;

Qy 122 DOEVIKASNSKIRLEKGLYQI-KIYOAREN-----PREKGLDFL----- 163
Db 202 DIIINFLSNRYK-----YDIGKLNQKKNKNNKMAIGATIDINTENILTNKLKSY 253
Qy 164 YWTDSONKKEVISSDNLQIPELKOKSSNSRKRSTAGPTVPDRNDGIPOSLE-VEGYT 222
Db 254 YQELIOTNNLAASDILRIALIKKYG-----GYCOLDPLPGVN 292
Qy 223 YDVKKKRTFLSPWISN-----IHEKGLTKYKSSPEKMWSTADSPYSDFEKVYGR 272
Db 293 LSLFNDISKPNQMDSNYWEALFEALIANEKKLMNY-----PKYMEQVPEI 340
Qy 273 DKNVSPPEAR-----HPLVAAYPIVHDMENIILSKNEDOSTONTDEFTTISKNTS-TSRT 327
Db 341 KERILISFYRNHINDLILPLGDIKISOLEILSLRKAATGKTFENAFIISNDSLTINN 400
Qy 328 HTSEVHG-----NAEVHANTSTSRTHSEVGNAEVHAVALDHSLSLAGEFTMAETWC-L 381
Db 401 LISOLENRYEILINSIOEKFKICEFYDSYINSVSELYETTPKKNLSMGSSSEFYQOIIQVL 460
Qy 382 NTAADARLANANRY-----VNTGTAPIYVNLPTT-SLYGKNQOTLATIKAKENOSOLA 435
Db 461 SSGFRPEVNSIVFFSGPNYISSATCTYHFKNTDMLSSOQET-----FEA 508
Qy 436 PNNYPSKNLAPIALNAODDFESPTI--TNMYNOFLEKTKQLRLDPOYGNATYVN 492
Db 509 SNMLYFSK-----THDEFFKSWLNSNIAEKFEOKLIKTYIGR-----TLN 549
Qy 493 FENGR-----VRVDTSQMWSEVLPOIETTARIIFENGKDLNVERRIAVNPSPDLETTK 547
Db 550 YEDGILFNFKMKRYVT-----SELLKYIEEVNSTRKYENYDLMNI-----LQIOG 593
Qy 548 PDMTLKEALKIAFGNEPKNQYOGKDIETEDFNDQ--QTSOINKQOLAEALN----- 599
Db 594 DDISTESAVNV-FGKN-PKKSILIOGVDDFANVFYFENGIVOSDININILSFNDIKRIK 651
Qy 600 -----ATNIYTVLDKIKLN-----AKMNI----- 619
Db 652 LTLIGHGENVFPKLFEGKQTVNDLTNNIKRQLHLLREGVILKNKYLKINILDCMYFT 711
Qy 620 -----IRDKRFH-----YDRNNNAVGADSESV--KEAREVINSSTEGLL 658
Db 712 PKVDINSTFVGKLFENKISRDLQPKGFSKNOLEISANKYAIRINREGCKREVLDYFGK-AWS 770
Qy 659 NIDKDIRKILSGYIV--EIEDT-----EGLEKY-----INDRDMLNISLRD-QKT 703
Db 771 NTDLIABOISNRYVYVWMEVENTLSARVEQLKKVAEPAKDINSITIOTTNDELKOSLVNT 830
Qy 704 FIDF-----KKYNDKLPLYISNPYKVVAVTKE-----NTIINSENGDSTTNG 749
Db 831 YADLITFLYSELKEDIPELDNIQIKRI--ILNEISRLHDFSNIILDFQKKNISNM 888
Qy 750 IKKIL-----IFSCKGY 761
Db 889 I--ILFDSIIEKEDY 901

RESULT 11
```


QY 611 KLNAAKNIILRDKRPHYDRN-----NIANGADESVYKAEHRE-----VI----- 649
Db 1621 -----VVKYKDIGIHYDCKRKKVIAITDASVGEYDAOSRIGVILMYGMNIPNVYS 1673
QY 650 NSTEGLLINIDDKIRKISGY-----IVETEDP-----EGKK----- 682
Db 1674 NKSTNACVSTELALHETEGYADSETLKVYTLKELEGCGNDNDIWTDSKPAIQGLNRSY 1733
QY 683 -----EVINDRYDMLNISSLRDQKTFI-----DFKKY 710
Db 1734 QQPKKFTWIKTEIKERIKESIKLKLTGKGNINADLLTKPVSADSKFRF 1784
RESULT 13
568218
botulinum neurotoxin type A, nontoxic/nonhemagglutinating - Clostridium botulinum (strain
C:Species: Clostridium botulinum
A:Variety: strain NIH
C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 15-Oct-1999
C:Accession: 568218; 574301
R:Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
PDBS Lett. 376, 41-44, 1995
A:Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components
A:Reference number: 567988; MUID:96096783
A:Accession: 568218
A:Molecule type: DNA
A:Residues: 1-1193 <FUJ>
A:Cross-references: EMBL:D67030; NID:g2160224; PIDN:BAAL1050.1; PID:d1011710; PID:g11324
A:Experimental source: strain NIH
A:Accession: 574301
A:Molecule type: protein
A:Residues: 1-13; 145-155 <FUJ>
A:Experimental source: strain NIH
C:Genetics:
A:Gene: ant
C:Keywords: neurotoxin
F:1-1193/Product: botulinum neurotoxin type A, nontoxic/nonhemagglutinating #status exp

Query Match 4.7%; Score 182.5; DB 2; Length 1193;
Best Local Similarity 18.6%; Pred. No. 0.12;
Matches 187; Conservative 144; Mismatches 351; Indels 323; Gaps 47;

QY 14 STILVSTGNELVIAQAEKQENRLNSESSSQGL-----LGYFSDLNFOQPM 62
Db 167 SNIVIFPG-----ANIYENTVFKKEDANGMGTMEIWFQPLFYKYDEFYIDPAI 220
QY 63 VYTS-----STYGDLSTP-----SSELENIPEMQYQSAIWSGFIVK--KSPD 105
Db 221 ELIKCLILSLFLYGIKRPDDLVIPRLRSELENIYSQLNIVDLLVSGIDPKFTINTDP 280
QY 106 YTFATS-----ADNHTVMVADQEVINKASNSMKILER-----GLYQIKIQY-- 149
Db 281 YWFTDNYNSNAKKVPEDHNIYETELEGNNALGNDIKLRLKQKFRNTINDIMELNLYTS 340
QY 150 -----QREPTKEGLDFKLYW-----TDSQNKKE 173
Db 341 KEFSIMDPDRFNNAKHFYKQYKIDYPENYSINGFVNGQJNLAQLSLSDRMQDIIINKPE 400
QY 174 VTSS-----DNQLP-----ELKQSSNSKKKSTSG 201
Db 401 ETLINLNGNNVSLMSNITYGDLKSTVDDEFYSNYKIPYRAVEYHFNNSNDSSLDNVNG 460
QY 202 PTVPRDNDGIPDSLEVEGYVD-----VKNKRFTLS--PAISNIHEKKGLTKYKS 250
Db 461 VI-----DNIPEIIDVNPYKENCOKESPVKITSTREINTIIPWPIV--LQAQNT 509
QY 251 SPEKSTASDPYSDPEKYTGRIKDNVSPARHPVLAAYPIVHDMENII--LSKNEDOST 308
Db 510 NNEKFSLS-----SDEVEYVSSKDKS-----LVYSFLSNVWFYLDISKDNP 551
QY 309 QNTDSE-----TFTISKNTSTSRTHTSEVHGNAEVAHANT-----STSRHTSEVHG 354

Db 552 IDTKKYYVLMLEIFRNFYSFDITATQEIINTGCKINKVVTWFGKALNLTSDSFVEEFQ 611
QY 355 NAEVHVALIDHSLSLAGERTWA---ETMGLNADPFRLANRNVYTG--TAPIVNVLP 409
Db 612 LGAISLJNKKENISMPITSIEYLPNDMLPLNDLNEKLPNIYSKNTAVFKKITYINFLDQ 671
QY 410 -----TSLVICKNOTLA--TIKAKENQLSOILAPNNYPSKNLAPIALNA--- 452
Db 672 WMTQYYSQYFDLICMAKRSVLAQETLIKRIIQKLSYLLGNSNISDNLALNLTPTNTL 731
QY 453 QDDFSSPTITM-----YNOFLEKTKQLRDTDOYG--NIATYN 492
Db 732 RDISNESQIANNVDSFLNMAAICVFESNIYKFIISF-----MEQCINNINIKTKE 782
QY 493 FENGRRVVTGSMSEVLPOIETTARIIFNGKDLNLVERIAAVNPDPLETPRDMTL 552
Db 783 F-----IQKTNINDEKLOLINQ---VNSLDEFFLN-----IQNKSLSFSETALIT 829
QY 553 KEAL-----KIAFGFNEPNCN-----LQYQKDI--TEFDNFD-----OOTSON 590
Db 830 KEETWPELVLYAFKEPQNNVIGDASGKNTSIEY--SKDIGLYGINSDALYLSNGSIS 888
QY 591 INQLAELNATNITYV-----LDKIKLNAM----- 616
Db 889 FSNDEFENGLTNSFSIYFWLRNIGDKTISKLSKEDNCGEIIFYODTGLVFNMIDSNG 948
QY 617 ---NILIRD---KRPY-----DR--NNIANGADESVV--KEAHREYVSTEGLLINDK 662
Db 949 NEKNITLSVSNNSMHTYITSDRLKEQLLITIDDLVLANESIKELNITYSNITSLISE 1008
QY 663 DIRKLISGYIEIEDTEGLEVINDRYDMLNITSLRQDQKTEIDFKKYNDKLPYISMPN 722
Db 1009 NNPYSIEGLTILNKPTTS--QEVLSNFEVLNNSYIRDSNEERLE---YKTYQLV-----N 1060
QY 723 YKVN---VYAVTKENTI---INPSENGDTSTGKIKILIFSCKGY 761
Db 1061 YVFSOKPICEVKQNNNIYTLTNTNMLNLQASKFKLSTINPKQY 1105
RESULT 14
555101
hypothetical protein YMR219w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR261.13; hypothetical protein YMR9959.01
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 09-Mar-1996 #text_change 29-Oct-1999
A:Accession: 555101; 557587
R:Bedman, K.; Brown, D.; Bowman, S.
submitted to the EMBL Data Library, June 1995
A:Reference number: 555089
A:Accession: 555101
A:Molecule type: DNA
A:Residues: 1-711 <DED>
A:Cross-references: EMBL:Z49809; NID:g854459; PIDN:CAA89934.1; PID:g854471; MIPS:YMR2
R:Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, June 1995
A:Reference number: 557587
A:Accession: 557587
A:Molecule type: DNA
A:Residues: 608-1658 <SKF>
A:Cross-references: EMBL:Z49939; NID:g887599; PIDN:CAA90190.1; PID:g887600; MIPS:YMR2
A:Experimental source: strain AB972
C:Genetics:
A:Map position: 13R
Query Match 4.7%; Score 182; DB 2; Length 1658;
Best Local Similarity 19.3%; Pred. No. 0.21;
Matches 161; Conservative 133; Mismatches 308; Indels 224; Gaps 39;

QY 25 EYIOAEVQENLNLNSES--SSQGLGYYFSDNLNQAAPVYVSSITGDSLSPSSLENIIP 83

Db 287 EGVMELEDDIDVESDAEKDESOGAEGTEHS -VDFSKYMQPTDNTKIPVIEKYESDEHK 345
QY 84 SEMQYFQSAIWS-GFIKKVKSDE-----YFATSADN---HYMMVVDQVINKASN 131
Db 346 VHQRYSEDGAFDFGSVNIYVDDESEDESESOAESYSAENAVYHNEHELDKELIEDIES 405
QY 132 SNKIRLEKGLYOIKIOYQRENTEKGLDFKL-----WTDSQNK 171
Db 406 SDS-----ESQASQESQSEDEDFEKKMNEKSTSETEENTSROQRAKADYTK 456
QY 172 KEVISSDNLOLPEL-----KQSSNSRKKRSTAGPTVPDRDNGIPDLS 216
Db 457 NKVEQOENDEEPEKDDIIRSSLDKNFHNKKEGYSENVLENETDPAIVERNO-IND-- 513
QY 217 EYEGYVVDVKNKRTFL--SPWISNIHE--KKGILTKKSS-----PEKWSADPY---S 263
Db 514 -VEGYVNTGKSVESDLHEHSP--DNLYDLAARMLQFOQSRNSNCPQKEQVSESYGHS 570
QY 264 DFEKVTGR-IDKNVSPEARHPLVAAYPIVHVMENIILSKNEDOSTQNTDSERTISKNT 322
Db 571 NGSNLISGRIDES---EQIPLKD-----FTGENNNIKTKTRGDLSSVEI 613
QY 323 STSRRTSEVHGNAEYH-ANTSTSRHTSEVHGNAEYHAVAIDHSLSLAGEERTWETGL 381
Db 614 EYEVSEKKIDGSTEKELVPLSTDTIINSSLGNEISYSLDDADAISENLTDVPLMEI 673
QY 382 NFA-----DTARLNMIRVYNTGTATVYVLPPTSILVGLKNQIAT 422
Db 674 KTPPKTEVVISSESVSTSEYEDNTVAMPQVEY---TSPFMD-PENSL---ND--D 721
QY 423 IKAKENQSLIAPNNYFSKNIAPIALNAODFSSPTIM-----NYNOFLELEKT 474
Db 722 YEKKHDLKSTLAA-----LAPAFKKDAFEVGEVTKSCULTISGHTNFHSKET 773
QY 475 KQRLDLDQVYGNIAIYFNENGHRVDTGSNMSEVL-PQIQTETARILFNGKDLNVERRI 534
Db 774 KOVS-DLDESTEVENYFENEWTG---DENKQSGKNFPGVANSSTDKSTEDNTD---EKYE 824
QY 535 AAVNPDPLETTKPDMTLEKALKIAGFNPNNGNLOYOGDITEPDPNF--DQOTSQ-- 590
Db 825 SAINVYN---VHGSSCEDIIETASVVE--NLRCEMDMEAEISSGDECVKQDGD 877
QY 591 IKNOQA-----ELNATNIYIVLDKIKLAKKNNIILRDKRFYDRNNIIVADESVYKE 643
Db 878 SKQIJSFSDSPNQESNDNTEFSSTK-----YKANSLEDEDESJKE 922
QY 644 AHR-EVIN-----SSTEGLLNIDKDIRKILSGYIIEDETEGLK 682
Db 923 LTKAEVYDKLDEESEDYBQDYADEPQNGDEGSNENIYVGTAK-----DTLGIV 972
QY 683 EYVNDYVDMINISLQDQKTFIDFKKYNDKLPYISNPKYKVVAVYKENTLIN 738
Db 973 EPENEKVN-----KVHEETLFEANVSSVAVQNMKDMTVDVIN 1010
RESULT 15
T18351
Impl protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 07-Dec-1999
C:Accession: T18351
R:Jensen, L.T.; ladefoged, S.; Birkelund, S.; Christensen, G.
A:Title: Selection of Mycoplasma hominis Pcz1 deletion mutants by cultivation in the pre
A:Reference number: Z18884; MUID:95369882
A:Accession: T18351
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1051 <JEN>
A:Cross-References: EMBL:U21961; NID:g790241; PID:g790242; PIDN:AAA81012.1
C:Genetics:
A:Gene: Impl
A:Genetic code: SGC3

Query Match 4.68; Score 180; DB 2; Length 1051;
Best Local Similarity 19.78; Pred. No. 0.14;
Matches 170; Conservative 137; Mismatches 324; Indels 232; Gaps 43;
QY 2 KRRVILPLMAL-----STILVSTGNGLEVIOAEVKOKNRLNSESSESSQ--GLGY- 51
Db 4 KKKRIALPLILCGGLAIATTAIAIKARRHKIAKANSQKERNLQNFKLQKLNELLGYK 63
QY 52 YFSDLN-FQAPVYVTSSTGDLSPS--SELENPSENOYFQSAIWSGFIKRV----- 101
Db 64 IYVEINVFHQEVL-----QGSILKINNKSETKAIEETLRLKDATITLLISKINGINQEL 119
QY 102 -----KSDRY---TATSADNHVYKVVDDQEVYINKASNKIRLEKGLYQIK-- 146
Db 120 EFAKFNEIKRLOEYIKNEISKOYEYHI-----KQNIENELNKVTPISLE-STLIEIOMA 173
QY 147 -----IOYQRENPEKGLDFKLYWTDSONKKE---VSSDNLOLPELKQSSNSRKKRST 198
Db 174 TNNILKILNISTEKKN-----IDNLNAKEQLKASISQANOLLPQSLDNDSEIATKAKK- 226
QY 199 SAGPTVPDRDNDGIPDSLEVE-----GYVDVKNKRTFLSPWISNIHEKKGLTKY 248
Db 227 -----SLDAEIKKANQAVASNNMTASQSAKSSLDAAKVAETTK--LETf 268
QY 249 KSPPEKWSADPYSDPEKYTGRIKAVSPEARHP---LYAAYPIVHVMENIILSKNE 304
Db 269 NKDKEA-----KFNEIKQNRNDIOEFINTKNPNYSELISQITSKR-DKSNVYDSSN 321
QY 305 DOSTQNTDSEFRTISKNTSRTHTSEVHGNAEYHANTSTSRHTSEVHGNAEYHAVAID 364
Db 322 KSDIESANTELKQALAKANDKQVADNLAKSIKQLNNSVSNANT-----LSALTD 373
QY 365 HSLSLAGERTWAE-----TGLNTADTARLNMIRVYNTGTATVYVLPPTSILVGLKNQ 418
Db 374 KDMTIOAKATELEKEIOKANQAIKSNNTASQSAKSSLDAAKVAETTKLETFN--KQKEA 431
QY 419 TLATIKAKENQSLIAPNNIYF-----SKNLAPIALNAODFSSPTIMN 464
Db 432 KFMELKOTRQIOEFINTKNPNYSELISQITSKRDSKSNVYDSSNKSIDEA--NTE 488
QY 465 YNOFLELEKTQRLDLD--QVYGNIAIYFNENGHRVDTGSNMSEVL-PQIQTETARIIIF 522
Db 489 LKQALNTAKAKSSIDNLPKLNQDSKIEEGPIR-NTNFSWT----- 532
QY 523 NGKDLNLYERRIAAVNPDPLETTK---PDMTLKALKIAGFNEPENG-- 568
Db 533 -----SSKLETTKKNLAEIETKADAIK-----NPNSSSQKQALKDSSQ 570
QY 569 LQYOG---KQITEPDPNPOQOTSQNIKNQDLAELNATNIV--TVLDKIK-----LN 613
Db 571 VQKIGNELKTITE-EPGKVEYTKNSNIGYRLFKLAQEQCRNNSVDPLKNAWEKQTLIS 629
QY 614 AKNNIILRDKRFYDRNNIIVAGADESVYK-----EAH-REVINSSTEGLLNIDKDIRK 666
Db 630 KQKRLQNGSTKDYLTQJSTKSTQESTIKKYIVNIQAHINNNLSOYR---LEADKLIAN 686
QY 667 ILSGY-IYVEIETEGLEKVINDRYDMLNI-SSLRQDGKTFIDFKKYNDKLP---YIS 719
Db 687 MKRGYGGKVGIESLQKQDMLMD--SVLSVDDSLAKD-----FNKALRYLYGVGYTK 735
QY 720 NPKYKVVAVYKENTINPSEN 742
Db 736 NP--PVSSMFPINRNRSIENYQN 756

Search completed: December 2, 2001, 13:51:18
Job time: 313 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:48:22 ; Search time 112.1 Seconds
(without alignments)
504.835 Million cell updates/sec

Title: US-09-747-521-4
Perfect score: 3913
Sequence: 1 MKRRKVLPLMALSTILVSS.....TSTNGIKKILFSKKGEYIG 764

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A.Geneseq_1101.*
2: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.*
3: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.*
4: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.*
5: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.*
6: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.*
7: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.*
8: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.*
9: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.*
10: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.*
11: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.*
12: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.*
13: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.*
14: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.*
15: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.*
16: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.*
17: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.*
18: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.*
19: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.*
20: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.*
21: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.*
22: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.*
23: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3800	97.1	764	21	AAV56958
2	3800	97.1	764	22	AAW47306
3	3699.5	94.5	763	21	AAV56960
4	3667	93.7	735	15	AAW60179
5	3667	93.7	736	21	AAV56959
6	3622.5	92.6	903	15	AAW60183
7	3490	89.2	719	15	AAW60193
8	2812	71.9	569	21	AAV56961
9	790.5	20.2	884	17	AAW91239
10	789.5	20.2	884	17	AAW19509
11	789.5	20.2	884	18	AAW46712

12	789.5	20.2	1346	17	AAW91245	VIP2A(a) and VIP1A
13	789.5	20.2	1346	18	AAW19513	B. cereus VIP1A(a)
14	789.5	20.2	1346	18	AAW46723	VIP1A(a)/VIP2A(a)
15	787.5	20.1	880	21	AAW60224	Bacillus thuringie
16	787	20.1	881	21	AAV59277	MIS toxin from B.
17	784.5	20.0	884	15	AAW63793	Bacillus cereus 10
18	777	19.9	1338	17	AAW91247	VIP2A(a)-VIP1A(a)
19	777	19.9	1338	18	AAW19520	Maize optimised-B.
20	777	19.9	1338	19	AAW46731	VIP2A(a)/VIP1A(a)
21	776.5	19.8	784	21	AAW82948	MIS toxin of Bacil
22	775	19.8	852	17	AAW91246	VIP1A(a) protein w
23	775	19.8	852	18	AAW19516	Maize optimised-B.
24	775	19.8	852	19	AAW46727	Maize optimised VI
25	772	19.7	860	21	AAV59282	MIS-8 toxin from B
26	766	19.6	834	17	AAW91242	B. thuringiensis V
27	766	19.6	834	18	AAW19512	B. thuringiensis V
28	766	19.6	834	19	AAW46722	Vegetative insect
29	596.5	15.2	425	19	AAW60222	Bacillus thuringie
30	587	15.0	667	17	AAW91240	B. cereus VIP1 pro
31	586	15.0	667	18	AAW19510	B. cereus 80 kd VI
32	586	15.0	667	19	AAW46713	80 kDa VIP1A(a) to
33	584	14.9	667	15	AAW63794	Bacillus cereus 80
34	541	13.8	357	19	AAW60228	Bacillus thuringie
35	538	13.7	357	19	AAW60231	Bacillus thuringie
36	521	13.3	357	19	AAW60218	Bacillus thuringie
37	521	13.3	357	19	AAW60226	Bacillus thuringie
38	503	12.9	357	19	AAW60237	Bacillus thuringie
39	450	11.5	327	19	AAW60220	Bacillus thuringie
40	447	11.4	327	19	AAW60223	Bacillus thuringie
41	427.5	10.9	348	19	AAW60219	Bacillus thuringie
42	425.5	10.9	348	19	AAW60229	Bacillus thuringie
43	419.5	10.7	347	21	AAW60221	Bacillus thuringie
44	419.5	10.7	347	21	AAV59276	Toxin from B. thur
45	396	10.1	333	19	AAW60230	Bacillus thuringie

ALIGNMENTS

RESULT 1
ID AAV56958 standard; protein: 764 AA.
AC AAV56958;
XX
DT 25-APR-2000 (first entry)
XX
DE B. anthracis protective antigen (PA) protein.
XX
KW Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;
KW tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.
XX
OS Bacillus anthracis.
XX
PN WO200002522-A2.
XX
PD 20-JAN-2000.
XX
PE 09-JUL-1999; 99WO-US15568.
XX
PR 10-JUL-1998; 98US-0092416.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;
XX WPI: 2000-182165/16.
XX DR N-PSDB: AAZ56874.
XX
PT Recombinant DNA construct useful as vaccines for anthrax, in producing
PT host cells for analyzing the drugs and agents inhibiting anthrax
XX
PS Disclosure; Page 33; 35pp; English.

XX The invention provides a recombinant DNA construct that comprises a
 CC vector and at least one nucleic acid (or its fragment) encoding a
 CC combination of *Bacillus anthracis* proteins, selected from protective
 CC antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
 CC (PA with its secretory signals replaced with those of tissue plasminogen
 CC activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
 CC for anthrax and in producing infectious alpha virus particles. These
 CC particles, expressing the B. anthracis proteins are useful also as
 CC vaccines for anthrax. Host cells transformed with the construct are
 CC useful for analyzing the effectiveness of drugs and agents that inhibit
 CC anthrax or B. anthracis proteins. The present sequence represents a
 CC B. anthracis PA protein.

XX Sequence 764 AA:

Query Match 97.1%; Score 3800; DB 21; Length 764;
 Best Local Similarity 97.6%; Pred. No. 2.4e-237;
 Matches 746; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

1 MKRRKVLPLMALSTILVSTGNLEVIOAEVKOENRLNSESSESSQGLIGYFSDLNFOA 60
 1 mkkrrvllplmalstlilvssstgnleivgaevkqenrlinesessqgllgyyfsdlnfqa 60
 61 PMVYTSSTTGDLSPSSSELENIPEENQYFQSAIWSGFIKKKSDERTPATSDNHVTMV 120
 61 pmvyssttgdlspssselempiensenyfqsaiwsgfikvksderytatsadnhvtmv 120
 121 DDQEVINKASNSNKRIRLEKGRLOYKIOYORENPTEKGLDFLYMTDSQNKKEVSSDNL 180
 121 ddqevinkasnsnkrlrlekgrloykirygrempetekgldfilywdsqnkkevssdnl 180
 181 QLPBLKQKSSNSRRKRSSTAGPTVPDRNDGIPDSLEVEGYTVDVKNKRTPLSPWISNTH 240
 181 qlpelkqkssnsrrkrsstagsptvpdrndgipdslevegtyvdvknkrtplspwismth 240
 241 EKKGLTKRKSSPEKSTSDPSDFEKYTGRIKDNVSEPARPLVAAPIYAVDMENITL 300
 241 ekkgltkryksspekstsdpsdfekytrgldknvseparplvaapyiyavdmennitl 300
 301 SKNEQSQNTDSEPTRTISKNTSTRTHTSEVHGNAEVAHANSTRTHTSEVHGNAEVAHA 360
 301 skneqsqntdseprtiskntsttrhtsevhgnaevahansttrhtsevhgnaevaha 360
 361 VAIDHSLSLAGEERTWAEFTWGLNTADTARLANIRYVNTGTADIVNLPTTSLVLGKNQTL 420
 361 valdhsislagerwaelmglnadtarlaniryvntgtadivnlpttslvlgknqtl 420
 421 ATIRKKNOLSOITLAPNNVYPSKNLAPALNADDESSPTTMANNOPLELEKTQOLRID 480
 421 atirkknolsgitlapnnvypsknlapalnadddesspttmannyqfielektqlrid 480
 481 TDQVGNATATYFENGRRVVDTSNMSEVLPOIQETTARIIFNGDLNIVERRIAAVNS 540
 481 tdqvgnatatyfengrrvvdtsnmsevlpoiqettariifngdlniverriaavns 540
 541 DPLETTKADMTLKEALKAFGEFNEPENGULQYQKNDITERDFENFDQOTSQNIKNQLAELNA 600
 541 dplettkadmtlkealkafgefnepengulqyqknditerdfenfdqotsqnknqlaelna 600
 601 TNIYIVLDKIKLNAMKNLIRDKRHHYDRNNIYAVGADESIVVEAAREVYINSTEGLLNTI 660
 601 tniyivldkiklnamknlirdkrhhydrrnniyavgaadesivveaarevyinstegllnti 660
 661 DKDIRKISGYIVIEDEPTEGLKEVINDRYDMLNISLSRODGKTFDFKYNKPLPLYTSN 720
 661 dkdirkisgyiveiedepTEGLKEVINDRYDMLNISLSRODGKTFDFKYNKPLPLYTSN 720
 721 PNYKVVAVYAKRENTIINPSENGDSTNGIKKILFSKKGYEIG 764
 721 pnykvvavyakrentiinpsengdstngikkilfskkyeig 764

RESULT 2
 AAB47306
 ID AAB47306 standard; Protein; 764 AA.
 XX
 AC AAB47306;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 XX Wild type B. anthracis protective antigen.
 DE
 XX Lethal factor; LF; immunogen; LF4; protective antigen; PA; DNA vaccine;
 KW humoral; cell-mediated; immune memory response.
 XX
 OS *Bacillus anthracis*.
 XX
 FH Key location/Qualifiers
 FT Peptide 1..29
 FT /label= Signal peptide
 FT /note= "Not given in the specification"
 FT Protein 30..764
 FT /label= PA
 FT Peptide 204..764
 FT /label= pcPA
 XX
 PN WO200145639-A2.
 XX
 PD 28-JUN-2001.
 XX
 PE 21-DEC-2000; 2000WO-US34912.
 XX
 PR 22-DEC-1999; 99US-0171459.
 XX
 XX (OHIS) UNIV OHIO STATE RES FOUND.
 PA (GALT/) GALLOWAY D R.
 PA (MATE/) MATECZUN A J.
 XX
 PI Galloway DR, Mateczun AJ;
 PI WPI; 2001-408540/43.
 DR N-PSDB; AAC86016.
 XX
 PT Protecting animal against lethal infection with *Bacillus anthracis*, by
 PT administering wild type or mutated form of *Bacillus anthracis* lethal
 PT factor protein or its fragment or a nucleic acid encoding the mutated
 PT protein -
 XX
 PS Claim 5; Fig 2; 33pp; English.
 XX
 CC This sequence shows the B. anthracis protective antigen (PA). An
 CC immunogenic fragment of PA, pcPA, can be used to produce an immune
 CC response which protects an animal against lethal infection with
 CC *Bacillus anthracis*. DNA encoding the B. anthracis PA can be used
 CC in conjunction with DNA encoding the lethal factor (LF) in a
 CC DNA vaccine. Using a DNA vaccine which encodes the mutated LF protein
 CC or fragment alone or in combination with a DNA encoding the PA protein
 CC or its fragment, both components (humoral and cell-mediated) of the
 CC immune system are stimulated, which results in longer term immune
 CC memory response. The combined use of a mutated LF and PA gene or their
 CC fragments results in a higher level of immune response, as judged by
 CC overall serum antibody titers for LF and PA antigens, than the use of
 CC either LF or PA genes in separate immunizations.
 XX
 SO Sequence 764 AA:

Query Match 97.1%; Score 3800; DB 22; Length 764;
 Best Local Similarity 97.6%; Pred. No. 2.4e-237;
 Matches 746; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

1 MKRRKVLPLMALSTILVSTGNLEVIOAEVKOENRLNSESSESSQGLIGYFSDLNFOA 60
 1 mkkrrvllplmalstlilvssstgnleivgaevkqenrlinesessqgllgyyfsdlnfqa 60

```

QY 61 PMVTSSTGDLSPSELENIPSENOYFOSAIMSGFIKKKSEDEYFATSADNHVTMV 120
XX |||||
PI 61 pmvtsstgdlspseleenisengyfgsalsgflkkskdeyfatadnhvtmv 120
XX |||||
QY 121 DDOEVINKASNNKIRLEKGRLYQIKIYORENPTKEGDLFKLYWDSQNKKEVISSDNL 180
XX |||||
DB 121 ddeevinkasnnkirkirlekgrlyqikiygrenptekgldfklywdsqnkkevisdnl 180
XX |||||
QY 181 QPELKOKSSNSRKRKSTSGPTVPDRNDGIPDSLEVEGYTDVKKRFFLSWISNIH 240
XX |||||
DB 181 qpelkoksnsrkrkstsgptvpdrndgipdslevegtydvkkrfflspwlsnh 240
XX |||||
QY 241 EKKGLFKKSPKRWSTASPDYSDFEKVTGRIRKNSPEARHPVLAAYPIVHDMENITL 300
XX |||||
DB 241 ekkglfkkspekswstasdpdydfekvgrirknvspearhplvaaypivhdmennll 300
XX |||||
QY 301 SKNEDOSTONTDESETRTISKNTSTSRHTSEVHGNAEVHANTSTSRHTSEVHGNAEVHA 360
XX |||||
DB 301 sknedgstontdsetrtiskntstsrhtsevhgnaevhantstsrhtsevhgnaevha 360
XX |||||
QY 361 VALDHSLSLAGERTMAETMGLNADTARLANRIRYVNTGAPRYNVLPPTSLVGLKQNTL 420
XX |||||
DB 361 valdhsislagertmaetmglncdrtarlanriryvntgapiynvlpptslvlgkntl 420
XX |||||
QY 421 ATIKAKENOLSOILAPNNYPSKNLAPIALNADODFSPTTMMYNOFLELEKTOLRLD 480
XX |||||
DB 421 atikakengsqllapnnypsknlapialnaoddfsspttmmynoflelektqlrld 480
XX |||||
QY 481 TDQVYGNATYFNEENGRVAVDTGSNNSEVLPOIETTARIIFNGKDLNVERRIAANPS 540
XX |||||
DB 481 tdqvygnatynfneengrvavdtgsnnsevlpoiettariifngkdlnverriaanps 540
XX |||||
QY 541 DPLETRKPMTLKALKIAGFNEPENGNOYOGKDTIEFPNPDQOTSONIKQOLALNA 600
XX |||||
DB 541 dpletckpmtlkaalkiafgnepngnlyqgkdtiefnfdqgtsqnlkqlaelna 600
XX |||||
QY 601 TNYTVLDRKIKLAKNMNIIIRDRKFHYDRNNIIVGADSVVKAHREVINSSTEGLLNI 660
XX |||||
DB 601 tnytvldrkiklaknmniiirdrkfhydrnniivgadsvvkaahrevinsstegllni 660
XX |||||
QY 661 DKDIRKILSGYIEIEDTEGLKEVINDRYDMLNSSLRODGKTFIDKKYNDKLLPYISN 720
XX |||||
DB 661 dkdirkilsygieiedteglkevindrlydmlnsslrodgktfidckkynndkplyisn 720
XX |||||
QY 721 PNKVVAVYATKENTIIINPSENGDTSTNGIKKILIFSKKGYEIG 764
XX |||||
DB 721 pnkvvavayatkentiiinpsengdstngiklilifskkyeig 764
XX |||||

```

RESULT 3

ID AAY56960 standard; protein; 763 AA.

AC AAY56960;

DT 25-APR-2000 (first entry)

DE B. anthracis TPA-PA protein.

KM Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA; tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.

OS Bacillus anthracis.

PN WO200002522-A2.

PD 20-JAN-2000.

PF 09-JUL-1999; 99WO-US15568.

PR 10-JUL-1998; 98US-0092416.

XX

```

PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL.
XX
DR WPI; 2000-182165/16.
XX
PT Recombinant DNA construct useful as vaccines for anthrax, in producing
XX host cells for analyzing the drugs and agents inhibiting anthrax
XX
PS Disclosure; Page 32; 35pp; English.
XX
CC The invention provides a recombinant DNA construct that comprises a
XX vector and at least one nucleic acid (or its fragment) encoding a
XX combination of Bacillus anthracis proteins, selected from protective
XX antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
XX (PA with its secretory signals replaced with those of tissue plasminogen
XX activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
XX for anthrax and in producing infectious alpha virus particles. These
XX particles, expressing the B. anthracis proteins are useful also as
XX vaccines for anthrax. Host cells transformed with the construct are
XX useful for analyzing the effectiveness of drugs and agents that inhibit
XX anthrax or B. anthracis proteins. The present sequence represents a
XX B. anthracis TPA-PA protein.
XX
SQ Sequence 763 AA:
XX
QY 3 KRKVLIRLMASTLIVSTGMLVYIOAEVKQENLNLNSESOGGLGYRSDNLFCQPM 62
DB 5 krgicvlllvcgavfas---eviaevqenllnesessgllgyyfsdlnfqpm 61
XX |||||
QY 63 VTSSTGDLSPSELENIPSENOYFOSAIMSGFIKKKSEDEYFATSADNHVTMVDD 122
DB 62 vtsstgdlspseleenisengyfgsalsgflkkskdeyfatadnhvtmwydd 121
XX |||||
QY 123 QEVINKASNNKIRLEKGRLYQIKIYORENPTKEGDLFKLYWDSQNKKEVISSDNLQ 182
DB 122 qevinkasnnkirkirlekgrlyqikiygrenptekgldfklywdsqnkkevisdnlq 181
XX |||||
QY 183 PELKOKSSNSRKRKSTSGPTVPDRNDGIPDSLEVEGYTDVKKRFFLSWISNIHEK 242
DB 182 pelkoksnsrkrkstsgptvpdrndgipdslevegtydvkkrfflspwlsnh 241
XX |||||
QY 243 KGLTFKRSPEKRWSTASPDYSDFEKVTGRIDKNNSPEARHPVLAAYPIVHDMENITLSK 302
DB 242 kgltfkrspekswstasdpdydfekvgridknnspearhplvaaypivhdmennllsk 301
XX |||||
QY 303 NEDOSTONTDESETRTISKNTSTSRHTSEVHGNAEVHANTSTSRHTSEVHGNAEVHANA 362
DB 302 nedgstontdsetrtiskntstsrhtsevhgnaevhantstsrhtsevhgnaevhana 361
XX |||||
QY 363 IDHSLSLAGERTMAETMGLNADTARLANRIRYVNTGAPRYNVLPPTSLVGLKQNTLAT 422
DB 362 idhsislagertmaetmglncdrtarlanriryvntgapiynvlpptslvlgkntlat 421
XX |||||
QY 423 IKAKENOLSOILAPNNYPSKNLAPIALNADODFSPTTMMYNOFLELEKTOLRLDPTD 482
DB 422 ikakengsqllapnnypsknlapialnaoddfsspttmmynoflelektqlrldtd 481
XX |||||
QY 483 QVYGNATYFNEENGRVAVDTGSNNSEVLPOIETTARIIFNGKDLNVERRIAANPSDP 542
DB 482 qvygnatynfneengrvavdtgsnnsevlpoiettariifngkdlnverriaanpsdp 541
XX |||||
QY 543 LETTRKPMTLKALKIAGFNEPENGNOYOGKDTIEFPNPDQOTSONIKQOLALNATN 602
DB 542 letckpmtlkaalkiafgnepngnlyqgkdtiefnfdqgtsqnlkqlaelnatn 601
XX |||||
QY 603 IYTVLDRKIKLAKNMNIIIRDRKFHYDRNNIIVGADSVVKAHREVINSSTEGLLNLNDK 662
XX |||||

```

|||||
Db 602 lvtvdtklknkmlllrdrfbydrnlaavadesvvrkaahrevlnssteglllnldk 661
Oy 663 DIRKLISGYIEIEDTEGLAKEVINDRYDMLNLSLRDQKTFIDPKKKYNDKLPYISMPN 722
Db 662 dirklisgyivreldegleikevindrmdlnslsrqdgkctfidkkyndkplpyismpn 721
Oy 723 YKVVAVYAVTKENTTIINPSENGDTSTNGIKKILIFSKGGEIG 764
Db 722 ykvvavvavtkenttiinpsengdstnglkkilifskgyeig 763

RESULT 4

AAR60179 standard; Protein: 735 AA.

AC AAR60179;
DT 03-APR-1995 (first entry)
DE Protective antigen of Bacillus anthracis.
KM Anthrax; Bacillus anthracis; fusion protein; protective antigen;
KW protective antigen; cell killing; targeting; targeting; pathogen;
KW Intracellular; HIV; human immunodeficiency virus; toxin.
OS Bacillus anthracis.
PN WO9418332-A.
PD 18-AUG-1994.
PE 14-FEB-1994; 94WO-US01624.
PR 12-FEB-1993; 93US-0021601.
PR 25-JUN-1993; 93US-0082849.
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX Atora N, Kimpel K, Leppla SH, Nichols PJ, Singh Y;
XX WPI: 1994-279753/34.
XX N-PSDB: AAQ70180.
XX Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumour cells
PT or HIV-infected cells
XX
PS Disclosure: Page 81-83; 124pp; English.
XX The sequence encoding the protective antigen of Bacillus anthracis
CC may be used in the construction of a nucleic acid which encodes a
CC fusion protein comprising the anthrax protective antigen binding
CC domain of the native anthrax lethal factor and a sequence encoding an
CC actively inducing domain of a second protein. The fusion proteins
CC are useful for the specific killing of tumour cells or the killing of
CC cells infected with intracellular pathogens, especially HIV.
XX
SQ Sequence 735 AA;

Query Match 93.7%; Score 3667; DB 15; Length 735;
Best Local Similarity 97.6%; Pred. No. 8.9e-229;
Matches 717; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

Oy 30 EYKQENRLLNESESSOGILGYFSDLNQAPMVVTSSTGDLSPSSLENIPEENOVF 89
Db 1 evkqenrllneesssgqllgyfisdlnqapmvvtssctgdlspsslelpsengyf 60
Oy 90 QSAIWSGEIKVKKSDYETFTSADNHVTMWVDDQEVINKASNSNKRIRLEKGRLYQIKIY 149
Db 61 qsaitsgfkvkkssdeyftatsadnhvtmwvddqevinkasnsnkrlrlekgrlyqiklqy 120

Oy 150 QRENTEKGLDFKLYWTDSONKKEYISSDNIQLPELKOKSSNSRKKRSTAGPIVPPDRDN 209
Db 121 qrenptekglfdklywtvdsqnnkkeyissdnlqlpelkksnsrkkrstesagpivppdrdn 180
Oy 210 DGIPTSLIEEGYTYDVKKRRTFLSPWISNHEKKGLTFYKSSPEKWSASDPYDPEKVT 269
Db 181 dgiptslieegytydvkkrtrflspwlnhekkgltyksspekwsasdpdydpekv 240
Oy 270 GRIDKNVSPPEARHPLVAAPYIVHVMENIILSKNEDOSTQNTDSETRTISKNTSTSRHT 329
Db 241 gridknvspearhplvaapylvhwmenilisknedgstqntdsetrtisknstsrht 300
Oy 330 SEVHGNAEVHANTSTSRHTSEVHGNAEVHVAVIDHSLSLAGEKRTMATGLNADTARL 389
Db 301 sevhnavevhasfidiagvsagfnsnsstvaldhsislageretwaelmglnadtarl 360
Oy 390 NANIRYVNTGTAPYNNVPTTSLYGKNOTLATIRAKENOSLILAPNNYPSKNLAPIA 449
Db 361 nanirvntgtaplynnvpttslygknotlatirakenglsqllapnnypsknlapia 420
Oy 450 LNAODPFSSPTITMNYNOFLELEKTKQLRLDQVYGNIAATYNENGRVRVDTGSNMSEV 509
Db 421 lnaoddfsstptitmnyngfielektkqlrldtqygniatyngfengrvrvdtgsnsew 480
Oy 510 LPQIDETTRARIIFNGKDLNLYBRRIAAYVPSDPLETTPDMTLKALKIAGFNPENGL 569
Db 481 lpqidettrariifngkdlnlvbrriaavpsdplettkpmtlkealkiagfnepgnl 540
Oy 570 QYOGKDIPEFPNFDQOSTSONIKNOELANTYTYVDKIKLNKMNILIRDRKFRHFR 629
Db 541 qyogkdipecfnfdqostsoniknglaelnatnlytvdklknkmlllrdrkfrfbydr 600
Oy 630 NNIAVGADESVYKKAHREVINSSTEGLLNLIDKDIRKLISGYIEIEDTEGLKEVINDRY 689
Db 601 nniavgadesvvrkaahrevlnssteglllnldkdirklisgyivledeglekevindr 660
Oy 690 DMLNLSLRDQKTFIDPKKKYNDKLPYISMPNPNVYVYATKENTTIINPSENGDTSTNG 749
Db 661 dmlnslsrqdgkctfidkkyndkplpyismpnykvvavtkenttiinpsengdstng 720
Oy 750 IKKILIFSKGGEIG 764
Db 721 ikkilifskgyeig 735

RESULT 5

AAY56959 standard; protein: 736 AA.

AAY56959;

AC AAY56959;
XX 25-APR-2000 (first entry)
DE B. anthracis MAT-PA protein.

XX Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;
KW tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.
XX
OS Bacillus anthracis.
XX
PN WO200002522-A2.

XX 20-JAN-2000.
XX 09-JUL-1999; 99WO-US15568.
XX 10-JUL-1998; 98US-0092416.

PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.

PI Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;

DR WPI: 2000-182165/16.

DR N-PSDB; AA556875.

XX Recombinant DNA construct useful as vaccines for anthrax, in producing
PT host cells for analyzing the drugs and agents inhibiting anthrax
XX
PS Disclosure; Page 34; 35pp; English.

XX The invention provides a recombinant DNA construct that comprises a
CC vector and at least one nucleic acid (or its fragment) encoding a
CC combination of Bacillus anthracis proteins, selected from protective
CC antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
CC (PA with its secretory signals replaced with those of tissue plasminogen
CC activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
CC for anthrax and in producing infectious alpha virus particles. These
CC particles, expressing the B. anthracis proteins are useful also as
CC vaccines for anthrax. Host cells transformed with the construct are
CC useful for analyzing the effectiveness of drugs and agents that inhibit
CC anthrax or B. anthracis proteins. The present sequence represents a
CC B. anthracis MAT-PA protein.

XX Sequence 736 AA:

Query Match 93.7%; Score 3667; DB 21; Length 736;
Best Local Similarity 97.6%; Pred. No. 9e-229;
Matches 717; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

```

OY 30 EVKQENRLNSESSESSGGLGYFSDINFOAPMVVTSSTGDLSPSELENIPSENOYF 89
    |||||||
DB 2 evkqenrlinesesssgllgyfisdlnfqapmvvtstgdlspseleinpsengyf 61
OY 90 QSAIWSGFIKVKKSDEYTFATSADNHVTMWVDDQEVINKASNSKIRLEKGRLYQIKIY 149
    |||||||
DB 62 qsaivsgfikvkksdeytfatsadnhvtmwvddqevinkasnskirkleqgrlyqikiy 121
OY 150 QRENPTKGLDFKLYWTDSONKKEVVISDNIQLPELKOKSSNSKRRKSTSGAPVPPDRN 209
    |||||||
DB 122 grenptekgldefklywtldsonkkekvisdnlqlpelkoksnskrkstsgapvpdrdn 181
OY 210 DGIPDSELENGYVVDVKNKRTFISPMISNHEKGLTKYKSSPEKWTASDPYSDEPKVT 269
    |||||||
DB 182 dgipdseleegyvdvknkrtfispmisnhekgltkyksespekwtasdpysdfekvt 241
OY 270 GRIDKNVSPRARPLVAAYPIVHVDMEITLSKNEDOSTONTSEPTITSKNSTSTRTHT 329
    |||||||
DB 242 gridknvspearplvaaypivhvdmeitlsknedostontseptitstksnststht 301
OY 330 SEVHGNAEVHANTSTSTRTHTSEVHGNAEVHAVAIDHSLSLAGERTMAETGLNTADTARL 389
    |||||||
DB 302 sevhgnaevhantststststststststststststststststststststststst 361
OY 390 NMIRRVNVTARIVNLPRTSLVGNKQTLATIKAKENLSQILAPNRYNPPSNLAPIA 449
    |||||||
DB 362 nmirrvnvtarivnlprrtslvgnkqtlatikakensqlilapnrypsnlapia 421
OY 450 LNAODESSPTPIPMANTNOPIELEKTKOLRDTQOVGNITVYENGRVAVDGSNMSEV 509
    |||||||
DB 422 lnaodessptpiimantnopielektkolrtdtovvgnitvynengrvavdgsnmsev 481
OY 510 LPOIETTARIIFNGDLNLVERRIAANVPSDPLETTKPMDTLKEAKIAFGFEPNGNL 569
    |||||||
DB 482 lpoiettariifngdlnlverriaavnsdpdlettkpmtlkealkiafgfepngnl 541
OY 570 QYQGNKITERDENFDQDTSNINKQALAEATNITVYLDIKILAKANNIILRDKRFHYDR 629
    |||||||
DB 542 qyqgnkiterdnfdqtsninkqalaeatnitvylldikilakannilirdkrfhydr 601
OY 630 NNIAVGADESVKAEHREVNSSTEGLLNIDDIRKILSGYVEIDTGLKRVINDRY 689
    |||||||
DB 602 nniavgadesvkeahrevinsstegllniddirkilsgyveidteglkevindr 661
OY 690 DMLNSSLRODGFTIDFKKYNNDKPLPLIYNPNKYVAVTKENTIINSENGDSTNG 749
    |||||||

```

DB 662 dmlnsslrgdqgkftidfkkyndkplyisnpnykvnvavtkeintiinsengdstng 721
OY 750 IKRLIFSKKGYENG 764
DB 722 ikrlifskkyeng 736

RESULT 6

AA60183
ID AAR60183 standard; Protein; 903 AA.

AC AAR60183;
XX

DT 04-APR-1995 (first entry)
XX

DE PA(1-725)-----Human CD4 fusion protein coding sequence.
XX

KW Anthrax; Bacillus anthracis; fusion protein; lethal factor;
KM protective antigen; cell killing; targeting; targeting; pathogen;
KW Intracellular; HIV; human immunodeficiency virus; toxin;
XX Pseudomonas; exotoxin.

XX Bacillus anthracis.
OS Homo sapiens.

OS Homo sapiens.
XX WO9418332-A.

PD 18-AUG-1994.
PF 14-FEB-1994; 94WO-US01624.

PR 12-FEB-1993; 93US-0021601.
PR 25-JUN-1993; 93US-0082849.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX

PI Atora N, Kimpel K, Leppla SH, Nichols PJ, Singh Y;
XX WPI: 1994-279753/34.

DR N-PSDB; AA070184.
XX

XX Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumour cells
PT or HIV-infected cells

XX Disclosure; Page 100-103; 124pp; English.
PS

XX This sequence is a fusion protein comprising amino acid residues
CC 1-725 of the anthrax protective antigen protein and residues 1-178
CC of human CD4, the portion which binds to gp120 on HIV infected
CC cells. Such fusion proteins may be useful for the specific killing
CC of tumour cells or the killing of cells infected with intracellular
CC pathogens, especially HIV, depending on their components.
XX

XX Sequence 903 AA:

Query Match 92.6%; Score 3622.5; DB 15; Length 903;
Best Local Similarity 96.7%; Pred. No. 9.1e-226;
Matches 710; Conservative 3; Mismatches 18; Indels 3; Gaps 1;

```

OY 30 EVKQENRLNSESSESSGGLGYFSDINFOAPMVVTSSTGDLSPSELENIPSENOYF 89
    |||||||
DB 1 evkqenrlinesesssgllgyfisdlnfqapmvvtstgdlspseleinpsengyf 60
OY 90 QSAIWSGFIKVKKSDEYTFATSADNHVTMWVDDQEVINKASNSKIRLEKGRLYQIKIY 149
    |||||||
DB 61 qsaivsgfikvkksdeytfatsadnhvtmwvddqevinkasnskirkleqgrlyqikiy 120
OY 150 QRENPTKGLDFKLYWTDSONKKEVVISDNIQLPELKOKSSNSKRRKSTSGAPVPPDRN 209
    |||||||
DB 121 grenptekgldefklywtldsonkkekvisdnlqlpelkoksnskrkstsgapvpdrdn 180

```

```

QY 210 DGIPDSLEVEGYTVVKNKRTFLSPWISNINHEKKGLTKYKSSPEKWSASDPYSPEKYT 269
Db 181 dgiplslevegtylvvknkrtflspwlsnlnhekgltkyspekwsasdpysdtekyt 240
QY 270 GRIDNVSPEARHPLVAAYPIVHVDMENTILSKNEDOSTOMTDETRTISKNTSRTHT 329
Db 241 gridnvspearhplvaayplvhwmentilsknedgstqntdsetrtlskntsstht 300
QY 330 SEVHGNAEVHANTSTSRHTSEVHGNAEVAHVAIDHSLSLAGERTAETMGLTADTARL 389
Db 301 sevhgnaevhasffdiggsvasgafsnstsvaidslslageretwaetmgintadtarl 360
QY 360 NANIRYVNTGAPRYNVLPPTSILVKGKNOTLATITAKENOSQLIAPNNYPSKMLAPIA 449
Db 361 nanirylvntgaprynvlpptsilvkgknotlatitakenqslilapnnypskmlapia 420
QY 450 LNAODFESTPTITMYNOFLERKTOLRLDPOYGNATYNEFNGRVRVDTSNWSRV 509
Db 421 lnaoddfestptitmyngflerktqlrltdqygnatynfengrvrvdtsnwsrv 480
QY 510 LPQIDETARIIFNGKDLNVERRIAANVPSDPLETTPKPDMLKEALKIAFGFNPNGNL 569
Db 481 lpgidettarilifngkdlnverriaavnsdpdlettkpdmllkealkiafgfnepngnl 540
QY 570 QYOGKDIREFDFNPDQOSTONIKNOLAELNATNTYTVDKIKLNKKMILIRDKRFHIDR 629
Db 541 qyogkdirefdnfdqtsqnlknglaelnatnlytvldkiklnkmlilrtdkrfhydr 600
QY 630 NNIAVGADEVYKREAHREYINNSTEGLLNIDKDIRKILSGYIVIEDEPLEGLEYINDRY 689
Db 601 nniavgadevvykrehreynssteglllnidkdlrklilsgyiveledteglekveindy 660
QY 690 DMLNLSLRQDGKTFIDFKKYNDKLPYISNPYKVNVAVTKEMTIINPSENGDTSNG 749
Db 661 dmlnlsrlrqdgktfidfkkyndkplyisnpnykvnyavtkemtilnpsengdstng 720
QY 750 IKKIL---IFSKKG 760
Db 721 ikkilkkvvlqkgk 734

```

RESULT 7
AAR60193 standard; Protein: 719 AA.

AC AAR60193:

DT 04-APR-1995 (first entry)

DE Modified protective antigen of Bacillus anthracis.

KW Anthrax; Bacillus anthracis; fusion protein; lethal factor;
protective antigen; cell killing; targeting; targeting; pathogen;
intracellular; HIV; human immunodeficiency virus; toxin.

OS Bacillus anthracis.

PN W09418332-A.

PD 18-AUG-1994.

PF 14-FEB-1994; 94WO-US01624.

PR 12-FEB-1993; 93US-0021601.

PR 25-JUN-1993; 93US-0082849.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Atora N, Klimpel K, Lepia SH, Nichols PJ, Singh Y;

DR WPI: 1994-279753/34.

DR N-PSDB: AAQ70189.

XX

```

PT Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumour cells
PT or HIV-infected cells
xx
xx
PS Example 6; Page 114-115; 124pp; English.
CC
CC The sequence encoding the protective antigen of Bacillus anthracis
CC may be used in the construction of a nucleic acid which encodes a
CC fusion protein comprising the anthrax protective antigen binding
CC domain of the native anthrax lethal factor and a sequence encoding an
CC activity inducing domain of a second protein. Such fusion proteins
CC are useful for the specific killing of tumour cells or the killing of
CC cells infected with intracellular pathogens, especially HIV,
CC depending on the second component. The protective antigen and other
CC toxins require proteolytic cleavage to acquire activity. Since some
CC cells infected with an intracellular pathogen possess an active
CC protease with quite a narrow substrate specificity e.g. HIV, the
CC protease cleavage site found in the native toxin is replaced with an
CC intracellular pathogen specific protease site (See AAR60184-88). The
CC protease in cells that are infected with an intracellular pathogen
CC cleaves the modified toxin which is then rendered active and kills
CC the cell. This sequence is a modified Bacillus anthracis protective
CC antigen which has the amino acids originally at positions 162-171
CC replaced with the HIV protease cleavable sequence described in
CC AAR60186.
CC
CC Sequence 719 AA;
SQ

```

```

Query Match 89.2%; Score 3490; DB 15; Length 719;
Best Local Similarity 93.1%; Pred. No. 2.4e-217;
Matches 688; Conservative 6; Mismatches 21; Indels 24; Gaps 2;

```

```

QY 30 EYKQENRLNESESSSGGLGYRFSDLNFOAPMVYTSSTGDISPSELENIPSENGYF 89
Db 1 evkqenrlneesssgglyrfsdlnfqpamvtsstgdispselenipsengyf 60
QY 90 QSAIWSGFIRYKKSDEYTPATSDADNHVTWVDDQEVINKASNSKIRLEKGLYQIKQY 149
Db 61 qsaIwsGFIRYKKSDEYTPATSDADNHVTWVDDQEVINKASNSKIRLEKGLYQIKQY 120
QY 150 QRENTEKGLDPKLTWTSQNKKEVYSSDNDLPELKKOKSSNS-----KKRSTSGRPVP 205
Db 121 qrenTEKGLDPKLTWTSQNKKEVYSSDNDLPELKKOKSSNS-----KKRSTSGRPVP 180
QY 206 DRDNDGIPDSLEVEGYTVVKNKRTFLSPWISNINHEKKGLTKYKSSPEKWSASDPYSD 265
Db 161 drdndgipdslevegtylvvknkrtflspwlsnlnhekgltkyspekwsasdpysdf 240
QY 266 EKVYGRIDKNVSPEARHPLVAAYPIVHVDMENTILSKNEDOSTOMTDETRTISKNTS 325
Db 241 ekvtgridknvspearhplvaayplvhwmentilsknedgstqntdsetrtlskntsts 300
QY 326 RHTSEVHGNAEVHANTSTSRHTSEVHGNAEVAHVAIDHSLSLAGERTAETMGLTAD 385
Db 301 rhtsevgngnaevhasffdiggsvasgafsnstsvaidslslageretwaetmgintad 360
QY 386 TARLANIRYVNTGAPRYNVLPPTSILVKGKNOTLATITAKENOSQLIAPNNYPSKML 445
Db 361 tarlanirylvntgaprynvlpptsilvkgknotlatitakenqslilapnnypskml 420
QY 446 APIALNAODFESTPTITMYNOFLERKTOLRLDPOYGNATYNEFNGRVRVDTSN 505
Db 421 apialnaoddfestptitmyngflerktqlrltdqygnatynfengrvrvdtsn 460
QY 506 WSEVLPQIDETARIIFNGKDLNVERRIAANVPSDPLETTPKPDMLKEALKIAFGFNP 565
Db 461 wsevlpgidettarilifngkdlnverriaavnsdpdlettkpdmllkealkiafgfne 520
QY 566 NGNLOYOGKDIREFDFNPDQOSTONIKNOLAELNATNTYTVDKIKLNKKMILIRDKRF 625
Db 521 ngnlqygkdirefdnfdqtsqnlknglaelnatnlytvldkiklnkmlilrtdkrf 580

```

Oy	626	HYDRNNINVAQESVYKREAHREVINSWTEGLLLNTIDKDIRKILSGYIVIEDETEGLKEYI	685
Db	581	HYDRNNIAVAGAESVYKREAHREVINSWTEGLLLNTIDKDIRKILSGYIVIEDETEGLKEYI	640
Oy	686	NDRYMLNINSSLRQGGKFEIDFKKRYNDKLPYISNPNTVYVYAAVTKENTINPSENGPT	745
Db	641	NDRYMLNINSSLRQGGKFEIDFKKRYNDKLPYISNPNTVYVYAAVTKENTINPSENGPT	700
Oy	746	STNGIKILILFESKGYEIG 764	
Db	701	stngikililfeskgyeig 719	
RESULT 8			
AY556961			
ID	AA556961	standard; protein; 569 AA.	
AC	AA556961		
XX			
XX	25-APR-2000	(first entry)	
DT			
XX			
DE	B. anthracis PA63 protein.		
XX			
KW	Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;		
KM	tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.		
XX			
OS	Bacillus anthracis.		
PN	MO200002522-A2.		
PD	20-JAN-2000.		
XX			
PE	09-JUL-1999; 99WO-US15568.		
XX			
PR	10-JUL-1998; 98US-0092416.		
XX			
PA	(USME-) US MEDICAL RES INST INFECTIOUS DISEASES.		
PI	Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;		
XX			
DR	WPI: 2000-182165/16.		
DR	N-PSDB: AA256877.		
PT	Recombinant DNA construct useful as vaccines for anthrax, in producing		
PT	host cells for analyzing the drugs and agents inhibiting anthrax -		
PS	Disclosure: Page 35; 35pp; English.		
XX			
CC	The invention provides a recombinant DNA construct that comprises a		
CC	vector and at least one nucleic acid (or its fragment) encoding a		
CC	combination of Bacillus anthracis proteins, selected from protective		
CC	antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA		
CC	(PA with its secretory signals replaced with those of tissue plasminogen		
CC	activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine		
CC	for anthrax and in producing infectious alpha virus particles. These		
CC	particles, expressing the B. anthracis proteins are useful also as		
CC	vaccines for anthrax. Host cells transformed with the construct are		
CC	useful for analyzing the effectiveness of drugs and agents that inhibit		
CC	anthrax or B. anthracis proteins. The present sequence represents a		
XX	B. anthracis PA63 protein.		
SO	Sequence 569 AA;		
Query Match 71.9%; Score 2812; DB 21; Length 569;			
Best Local Similarity 96.8%; Pred. No. 1,1e-173;			
Matches 550; Conservative 2; Mismatches 16; Indels 0; Gaps 0			
Oy	197	STSAAGPTVPDDRNDDIPDSLEVEGYTVDVKNKRTEFLSPWISNHEKKGLTKYKSSPEKMS	256
Db	2	stsaagptvpdrnddipdsleveytvdknkrftflspisnhkkglytkyxspekms 61	
Oy	257	TASDPYSDFEYKGTGRIDKNVSPDARHPPLVAAPVIVHVDMENTIILSKNEDOSTONTDESTR	316

Db	62	Tasopstefekvcgfridknvspeahpiivaayplivhgmellisknedgstqntdse	123
QY	317	TSKSTSTSRHTSEVHGNAEVEHANTSTSRTHTSEVHGNAEVHVAHDHSLSLAGE	376
Db	122	tksktstscrthtsevhnaevhasffldgsvsagfsmnsstvaldhsisigetw	188
QY	377	ETMGINTADTARLANIRRYVTGNAPRIVNVLPPTSLVLGKNQTLATIKAKENQS	436
Db	182	etmgintadarlanirryvtgnapriynvlpptslvlgkngtlatlakengsqll	241
QY	437	NNYYSKKNLAPALAAODFSSFTPTMTYNOPLELEKTKOURLDTDOYYGMIA	496
Db	242	nnyysknlaplalaaodfstptmtynyoflelektkqlrldtdygnlaeynfeg	303
QY	497	RVRVDTGSNMSEVFLPOIETTARIIFNGKDNLVERRIAANPSPDETTKPDML	556
Db	302	rrvvdtgsmsevrlpqietarrilfnfgkdnlverriaavnpdpdetcpmnlke	363
QY	557	KIAFGNEPNGLYOQGDKITEFDENFDQOTSQNIKNOLAELNATNTYTVDLIK	616
Db	362	kiafgnepnlglygqdfiefdnfdqtsnklngjaelnatnlytvldiklknkm	422
QY	617	NLLIDRKRFHIDRRNNIANGADEVYKKAHREYINSGLLNLNDKDIRKLISGY	676
Db	422	nllidrkrfhydrrnniavgadesvykkaehrevlnsstegllnldkdrklisgy	483
QY	677	DREGKEVINDRDYDLNISLRDCKTPIIDRKKYNDKRLPYISNPNKVVNYA	736
Db	482	dteglekeyindrdymnlsslrddgtklfdfkkyndkrlpyisnpnkvvnyavt	543
QY	737	INPSENGDPTSTNGIKKILIFSCKGEIG 764	
Db	542	inpsengdstngikkilifskgeig 569	
 RESULT 9 AAR91239 AAR91239 standard; Protein; 884 AA. AC AAR91239; XX DT 14-AUG-1996 (first entry) XX DE B. cereus VIP1A(a) insect-specific protein. XX KW Pesticide: Insecticide; biological control agent: Lepidoptera; KM Coleoptera: Transgenic plant; maize; Insect resistance; XX Western corn rootworm; Diabrotica virgifera virgifera: VIP. OS XX Bacillus cereus strain AB78 (NRRL-B-21058). XX FH Key FT Peptide Location/Qualifiers FT 1..33 FT /label= Sig_peptide W09610083-A1. XX PD 04-APR-1996. XX PF 27-SEP-1995; 95WO-EP03826. XX PR 05-JUN-1995; 95US-0463483. PR 28-SEP-1994; 94US-0314594. XX PA (CIBA) CIBA GEIGY AG. XX PI Carr B, Desai NM, Duck NB, Estruch JF, Kostichka K; PI Koziel MG, Mullinsma, Nye GJ, Warren GW; XX WPI: 1996-200921/20. DR N-PsDB; AAT13940. XX			


```
QY 110 TSADNHTMVDQEVINKASNSKIRLEKGRLYOIKROYORENPTEKGLD-----FKL 163
DB 117 lsedegallengllsnkqkqvnhlekqklyvlikleygsd--tkfnidskftkelkl 174
QY 164 YMTDSOKRKEVYSSDNLQLELKOKSS-----NSRKRSTSGAPTPDRDND 210
DB 175 fklsgnqpgqvqgdelpnfekkeqeflakpsklnltfcmkreided---tdcdgd 231
QY 211 GIPDSLEEGYTVDKNKRFLSPWISNIHEKGLTKRKSSPEKWSASDPYSDFEYVTG 270
DB 232 slpdlweengytl-----gnrlavkwddsl-askgytkfvsnpleshtvgpdytdekaar 286
QY 271 RIDKNVSPFARHPLVAAPYIVHVDMENTILSKNEDOSTQNTDSETRTISKNTSRTHTS 330
DB 287 dldlsnaketfnplvaafpsvnmekvllspnenls----- 323
QY 331 EVHGNAEVHANTSTSRHTSEVHGNAEVH-----AVAIDHSLSLAGERTWAETMG-- 380
DB 324 ---nsveshstmsylnntegasaevaglgpkqisfgysvnyghsetvage--wgtstgnt 378
QY 381 --LNTADTARLANIRVNTGTAPYINVLPTTSLVGLKNOTLATIKAKENQLOSLAPNN 438
DB 379 sqfntasaaylnanvynvgtgalydvkptsfvl-ndtlatlaktksnstaInispge 437
QY 439 YPFSKNLAPIALMAODESSTPTTMNNOFLLEKTKQLRLDPDQVYGNATYFENGVRV 498
DB 438 sykkkqngqglatlsmddfnshpiltlnkkqvdnllnmkpmletnqldg---vykikdthg 494
QY 499 RVDTSNMSEVLPOIOETTRIFNGKDLNIVERIAAVNPSPLETETKPDMTLEKALKI 558
DB 495 nltvgewngvlgqikaktasllvdgge-rvaektrvaekyempedct-psltckdskl 552
QY 559 AF--GNEPFGNLOYOQCKDITFERDF--NPDQOTSQNIKNOLAEL-----NATNIYVLDK 609
DB 553 sydelkelegllykxkpiyessvmltydenlaketvktjndctgkfkdvshlydv--- 609
QY 610 IKANAKNITLRDKRFHYDNNNTAVGADESVEKAHREVINSSSTREG-----LL 657
DB 610 -kltpkmnvtk-lsllydh---aesndnsigkwtcnlnvsgnngkqysnmnpdnl 664
QY 658 LNTD-----KDRIKILSGYVIEDTNE-----GKEVINDRYDMLN--- 693
DB 665 lntdageklknkrdyislmyksekntgcetldgelypiltktcvnknknykrldliah 724
QY 694 -----ISSLRQDGKTFIDKKYNDKLPYISNPKNVNYAYATKENTIIINPSENGDT-S 746
DB 725 nlsnpnlsslh-----lktndelctlfwdldsl-tdvasikpen--ltdseikqlys 772
QY 747 TNGIK--KILIFSKKGYEIG 764
DB 773 rygiklgedgillidckgqihyg 793

RESULT 11
AAM46712
ID AAM46712 standard; Protein; 884 AA.
XX
AC AAM46712;
XX
DT 02-JUN-1998 (first entry)
XX
DE 100 kDa VIP1A(a) toxin of Bacillus cereus.
XX
KW Vegetative insecticidal protein; VIP; expression; maize; protection;
KW plant; Ostrinia furnacalis; Asian Corn Borer; Cry toxin; VIP toxin;
KW recombinant; Bacillus thuringiensis; transgenic plant; resistance;
KW insect attack; Sesamia; maize; cereal crop.
XX
OS Bacillus cereus.
XX
PN M09746105-A1.
XX
```

```
PD 11-DEC-1997.
XX
XX 27-MAY-1997; 97MO-EP02737.
XX
XX 06-JUN-1996; 96GB-0011777.
XX
XX (NOVS ) NOVARTIS AG.
XX
PI Hunter B. Suwantaradon K. Utdewilligen WPM;
XX
XX WPI; 1998-041787/04.
XX
XX N-PSDB; AAV16166.
XX
PT Administration of Bacillus sp. toxin protein, especially Cry or
PT vegetative insecticidal protein (VIP) protein to plants - useful for
PT protection against attack by Asian Corn Borer (Ostrinia furnacalis)
XX
XX Claim 8; Pages 43-46; 175pp; English.
XX
XX The present sequence represents a 100 kDa vegetative insecticidal
XX protein 1A(a) (VIP1A(a)), and is derived from Bacillus cereus strain
XX AB78. The protein is used in a method for protecting plants and their
XX progeny against damage caused by Ostrinia furnacalis (Asian Corn Borer).
XX The protein is directly or indirectly applied to the plant, plant seed
XX or growing area of the plant. Cry toxins can also be used in the same
XX way, in place of VIP toxins. The Cry or VIP toxins and genes are used,
XX especially inside recombinant B. cereus or B. thuringiensis strains,
XX to produce plants protected against Asian Borer pests. Transgenic plants
XX protected against Asian Corn Borer can be used to produce seed and
XX Cry-type and a VIP toxin gene can also protect against Sesamia pests.
XX The method and compositions are especially used for protecting maize but
XX may also be used to protect other cereal crops against Asian Corn Borer
XX attack.
XX
XX Sequence 884 AA;
XX
XX
XX Query Match 20.2%; Score 789.5; DB 19; Length 884;
XX Best Local Similarity 29.4%; Pred. No. 1,le-42;
XX Matches 253; Conservative 134; Mismatches 309; Indels 165; Gaps 34;
QY 1 MKRRKVLIPLMALSTIIVSS--TGNLEVIOAEK-----QENRLNSESSESSQGLIGY 51
DB 1 mkmkkkkksavvtctllapmflngvnaayadskntgstqkng---qkemdtkqillyg 57
QY 52 YFSDLNFAQPNVNTSSTGDISIPSSLEN--IPSENOYFOSAIWSGFIYKKKSDERYFA 109
DB 58 yfkqkdf-snlmtfaptrdstllydqqtkankllidkkqgeysltwlglsqkctgdffln 116
QY 110 TSADNHTMVDQEVINKASNSKIRLEKGRLYOIKROYORENPTEKGLD-----FKL 163
DB 117 lsedegallengllsnkqkqvnhlekqklyvlikleygsd--tkfnidskftkelkl 174
QY 164 YMTDSOKRKEVYSSDNLQLELKOKSS-----NSRKRSTSGAPTPDRDND 210
DB 175 fklsgnqpgqvqgdelpnfekkeqeflakpsklnltfcmkreided---tdcdgd 231
QY 211 GIPDSLEEGYTVDKNKRFLSPWISNIHEKGLTKRKSSPEKWSASDPYSDFEYVTG 270
DB 232 slpdlweengytl-----gnrlavkwddsl-askgytkfvsnpleshtvgpdytdekaar 286
QY 271 RIDKNVSPFARHPLVAAPYIVHVDMENTILSKNEDOSTQNTDSETRTISKNTSRTHTS 330
DB 287 dldlsnaketfnplvaafpsvnmekvllspnenls----- 323
QY 331 EVHGNAEVHANTSTSRHTSEVHGNAEVH-----AVAIDHSLSLAGERTWAETMG-- 380
DB 324 ---nsveshstmsylnntegasaevaglgpkqisfgysvnyghsetvage--wgtstgnt 378
QY 381 --LNTADTARLANIRVNTGTAPYINVLPTTSLVGLKNOTLATIKAKENQLOSLAPNN 438
DB 379 sqfntasaaylnanvynvgtgalydvkptsfvl-ndtlatlaktksnstaInispge 437
```


PT vegetative insecticidal protein (VIP) protein to plants - useful for
 PT protection against attack by Asian Corn Borer (*Ostrinia furnacalis*)
 XX
 XX Claim 8; Pages 78-83; 175pp; English.
 PS
 CC The present sequence represents the fusion protein of vegetative
 CC insecticidal protein 1A(a) (VIP1A(a)) and VIP2A(a). The protein is
 CC used in a method for protecting plants and their progeny against
 CC damage caused by *Ostrinia furnacalis* (Asian Corn Borer). The
 CC protein is directly or indirectly applied to the plant, plant seed or
 CC growing area of the plant. Cry toxins can also be used in the same way,
 CC in place of VIP toxins. The Cry or VIP toxins and genes are used,
 CC especially inside recombinant *B. cereus* or *B. thuringiensis* strains, to
 CC produce plants protected against Asian Borer pests. Transgenic plants
 CC protected against Asian Corn Borer can be used to produce seed and
 CC progeny also resistant to insect attack. Plants expressing both a
 CC Cry-type and a VIP toxin gene can also protect against *Sesamia* pests.
 CC The method and compositions are especially used for protecting maize
 CC but may also be used to protect other cereal crops against Asian Corn
 CC Borer attack.
 CC
 XX
 SQ Sequence 1346 AA;
 Query Match 20.2%; Score 789.5; DB 19; Length 1346;
 Best Local Similarity 29.4%; Pred. No. 2,1e-42;
 Matches 253; Conservative 134; Mismatches 309; Indels 165; Gaps 34;
 QY 1 MKRKVLIPLMALSTIVSS--TGNLEVIQAEVK-----GENRLNSESSESSGGLG 51
 DB 463 mkmnmkklasvvtcllpmfingvnayadsktngistqknq---qkmdrkyllyg 519
 QY 52 YPSDLFQAPMVTSTTSDLSIPSELEN--IPSENYFQSAIWSGFIKVKSDRYTPA 109
 DB 520 yfkygxf-snlmfpdrstcllydqtanlkldkkgqgysirwglqsketgftfn 578
 QY 110 TSADNNTVMWVDQEVINKASNSNKIRLEKGRUYOKIKYOENPTEKELD-----FL 163
 DB 579 lsedegalliehgklsnmgkqkvnhlekylvpikleygsd--tkfidsktfkfekl 636
 QY 164 WYTDSONKKEVISDNQLPELKOKSS-----NSRKRKSTAGPTPPDRND 210
 DB 637 fklidsgnqgqygqdelrnpelnkkesgeflakpsklnlftqkmkreided---ldtgdg 693
 QY 211 GIPDSLEWEGYVVDKANKTFELSPWISNIEKKGLTKYSSPEKMSSTASDPYSDEKVTG 270
 DB 694 slpdlweengyftl---gnrlavkwddsl-askytkfivsnplshstlvdydyldekaar 748
 QY 271 RIDKNVSPARHPLVAAYPYIVHDMENILSKNEDOSTQNTDSERTIKNTSTSTHTS 330
 DB 749 dldltnakectfnplyaatpvsnvsmekvllspneils----- 785
 QY 331 EVHGAENVHANTSTRTHTSEVHGAENVH-----AAVAIDHSLAGERTWAETMG-- 380
 DB 786 ---naveshststnwstynlegasveaglsipklsfsvsnghseltvae--wgsctgnt 840
 QY 381 --LNTADTARLANIRYVNTGTAPIYVLPPTSLVLGKNQOTLATIKAKENQISQILLAPPN 438
 DB 841 sqfntasagylnaavrynmvgatgalydvktsfvl-ndtiatlakstalnspge 899
 QY 439 YPSKRLAPIALNADDESPRTIMNYNOFLELEKTKOIRLTDQVYGAIATYNEPENGAV 498
 DB 900 syppkkgqngialtsmdndtsnpltkkqvdnlmkkpmletnqgdg---yyklkldtng 956
 QY 499 RYDTGSNMSEVLPQIQTETARITFNCKDNLVERRIAAVNPDPLETTPTDWTLEKALKI 558
 DB 957 nlvtgngmngvlgkakatasiivddge-rvaektvaadgyenpedkt-psltlkdaiki 1014
 QY 559 AF--GFNEPNGMLQYQKADITEFDR--NFDQOSTONIKQJLAEL-----MATNITYTVLDK 609
 DB 1015 sypdsketkegillyknpkylesswmtlyidentakevtkqldntgkfkfdvshlydv--- 1071
 QY 610 IKLNKMKMLILIDKRFHIDRNNIIVAGADESVYKKAHREVINSTEB-----LL 657

DB 1072 -kltpkmwvlik-1sillydn---aesdnstgkwtntnlgvgnngkkyssnmpdanlt 1126
 QY 658 LNTD-----KDIRKILSGYVIEEDTE-----GLKRVINDRMDLN----- 693
 DB 1127 lntdagelknkndyilsymksektqceitldgelypitktvnnkndykrldilsh 1186
 QY 694 -----ISSLRDQKTFIDFKKYNDKRLPLYISNPKNVWVAVTKENTIIIPSENGDT-S 746
 DB 1187 nkspnlsish-----ikendeltlfdwdisl-tvasikpen--lttselkqys 1234
 QY 747 TNCIK---KILFSKKGTEIG 764
 DB 1235 rygikledgilldkkgilyng 1255
 RESULT 15
 AAW60224
 ID AAW60224 standard; Protein; 880 AA.
 XX
 AC AAW60224;
 XX
 DT 28-SEP-1998 (first entry)
 XX
 DE Bacillus thuringiensis insecticidal toxin 177C8.
 XX
 KW Insecticide; pesticide; toxin; delta-endotoxin;
 XX biological control; lepidopteran; coleopteran.
 FT Bacillus thuringiensis strain PS177C8 (NRRL B-21867).
 FT
 FT Misc-difference 675
 FT /note= "encoded by YTA"
 FT Misc-difference 846
 FT /note= "encoded by AC"
 FT Misc-difference 846
 FT /note= "encoded by RAA"
 PN W09818932-A2.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97WO-US19804.
 XX
 PR 30-OCT-1996; 96US-0029848.
 XX
 PA (MICO) MYCOGEN CORP.
 PI Dullum CJ, Feitelson JS, Loewer D, Muller-Cohn J;
 PI Narva KE, Schmelts JL, Schnepf HE, Schwab G, Stamp L;
 PI Stockhoff BA;
 DR WPI: 1998-272226/24.
 DR N-PSDB: AAV30307.
 XX
 PT Bacillus thuringiensis isolates - used for producing pesticidal
 PT toxins and nucleotide sequences for control of lepidopterans and
 PT coleopterans
 PS
 PS Claim 5; Page 81-84; 139pp; English.
 CC This polypeptide comprises a novel soluble toxin of *Bacillus*
 CC *thuringiensis* (B.t.) strain PS177C8 (NRRL B-21867). The toxin
 CC belongs to a novel family of B.t. toxins that have toxicity
 CC against non-mammalian pests. Its amino acid sequence was deduced
 CC from a novel DNA fragment (see AAV30307) obtained by PCR from
 CC cellular genomic DNA of PS177C8. Disclosed and claimed are novel
 CC B.t. isolates and toxins (see AAW60218-32) that have activity against
 CC lepidopteran and/or coleopteran pests, isolated genes, probes
 CC and primers (see AAV30288-321 and AAV9734-87) useful for production
 CC of the toxins and for the identification and characterisation of
 CC these toxins, and for the identification and characterisation of

CC bacterial hosts. The invention provides 8 entirely new families of
CC toxins from B.t. isolates. The toxins have the additional ability
CC to form pores in cell membranes, and can be used to facilitate
CC entry of a second agent into a target cell.

Search completed: December 2, 2001, 13:48:31
Job time: 151 sec

xx Sequence 880 AA;

Query Match 20.1%; Score 787.5; DB 19; Length 880;
Best Local Similarity 29.5%; Pred.No.1.5e-42;
Matches 251; Conservative 199; Mismatches 311; Indels 151; Gaps 35;

QY 1 MKRRKVLPLMALSTIVSS--TGNEVLEIAEVR-----GDNRLNESESSOGGLGY 51
DB 1 mkkk---lasvvtctllapmfingvnavyadsktngistctqknq---qkemdrtkglly 54
QY 52 YPSDLNFOAPMVVTSSTTGDLSPSELEN--IPSENOYPOSAIWGCFIKVKSDEYTPA 109
DB 55 yfkygkdf-snlmfaptdestllydqgtankliddkqegyslrwlglsgketgdfn 113
QY 110 TSADNHTVMWVDQEVINKASNSNKRIRLEKGRLYOIKIYQRENPTREKGD-----FKL 163
DB 114 lscdegaileingklisnkyekqyvhlekylvpikleyqsd--tkfnidsktfkcl 171
QY 164 YMTDSQNKKEVSSDNQLPELKOKSS-----NSRRKRRTSAGPTVPORDND 210
DB 172 fklsgnqpgqvgqdelmpelnfkkesgeflakpsklnltqkmkreided---tdltdgd 228
QY 211 GIPDSLEVEGYTVVKNKRTFLSPWISNIEKKGLTKYKSSPERKSTASDPYSDFEYTG 270
DB 229 slpdlweengytl-----gnrlavkwddsl-askyktcfvsnplshltvgdpytdyekaar 283
QY 271 RIDKNTSPEARHPLVAAPIYVHDMENILISKNDOSTONTDSETRTISKNTSRTHTS 330
DB 284 dldlsnaketfnpivaafpsvnmekvllspneils----- 320
QY 331 EVHGNAEVHANTSTSRTHTEVGNAEVH-----AVADHSLSLAGERTWAETWG-- 380
DB 321 ---nveshastnswytnntegasaevaglgpkglisfygsvnyghsetvage--wgtstgnt 375
QY 381 --LNTADTARLNNIRYVNTGTAPIYVNLPTSLVLGKNOQLATIKAKENQLSOILAPNN 438
DB 376 sgftasagylnanvynvgtgaiydvkptsfvl--nddtlatltaaknsalnispge 434
QY 439 YPFSKULAPIALNADDFSTPTTMNNOPLLEKTKQLRLDQVGNATATYFENGRLV 498
DB 435 syppkkgqnglatismddfnshpiltlnkkyvdnlmnpmmletnqtdg---vykkikdthg 491
QY 499 RVDYGSNMSEVFLPOIETTARIIFNGKDLNVERRIAAVNPSPLEETKPPDMTLKEALKI 558
DB 492 nlvtggevngvlgqakaktasiliyddge-rvaekryaakdyenpedkt-psltlkdaikl 549
QY 559 AF--GENEPNGNLQYOGKDIPEDF--NPDQTSQNIKNOLAEL----NATNIYTVLDK 609
DB 550 syppdelkelegllyknkpiyessvmltyidentakevtkgjndtltgkfkdvshlydy--- 606
QY 610 IKLNAAKNILIRKREHYDNNNIAVGADESVVAEAREVINSSTEG-----LL 657
DB 607 -kltpkmnyvlik-lsillydn--aesndnsigkwtlnclivsggnngkkyssnmpdanlt 661
QY 658 LNTD-----KDIRKILSGYIVETEDTE-----GLKEVINDRYDMLNI--SS 696
DB 662 lntdaqeklnknyislymksekntgcetltdgelypitlctvnnkndnykrlidilahn 721
QY 697 LRQDGKTFIDFKRYNDKPLIYISNPYKVVAVTKENTTIINPSENGDT-STNGIK--K 752
DB 722 lksnpisslnhtc-ndetltfwddisi-tdvasikpen--ltdselkqiyrsyglkiedg 777
QY 753 ILFSSKKGYEIG 764
DB 778 lildkkgghng 789

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:51:36 ; Search time 72.22 Seconds
(without alignments)
588.554 Million cell updates/sec

Title: US-09-747-521-4_COPY_178_735

Perfect score: 2871

Sequence: 1 DNQLPELKOKSSNRKRKRS.....LYISNPKNVAVATKENT 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2758	96.1	764	2	I39934	protective antigen
2	715.5	24.9	875	2	I40862	Iota toxin compone
3	211.5	7.4	192	2	I39933	cryptic protein -
4	211.5	7.4	204	2	G59104	hypothetical prote
5	181.5	6.3	2178	2	S55805	alpha-toxin - Clos
6	164.5	5.7	2529	2	B64635	toxin-like outer m
7	158	5.5	4688	2	F82885	hypothetical prote
8	157	5.5	1308	2	E71622	probable membrane
9	150.5	5.2	2269	2	T28677	thoptry protein -
10	150.5	5.2	2399	2	H71879	toxin-like outer m
11	150	5.2	1127	2	T28317	ORF MSV156 hypothe
12	150	5.2	1230	2	S56850	SMC1 protein homol
13	150	5.2	4919	2	T31105	hypothetical prote
14	149	5.2	1193	2	S68218	botulinum neurotox
15	148	5.2	810	2	D86818	hypothetical prote
16	147	5.1	1169	2	T18423	hypothetical prote
17	147	5.1	1939	2	T18372	repeat organelle
18	145	5.1	2401	2	T28676	thoptry protein -
19	144	5.0	2523	2	T18477	hypothetical prote
20	143	5.0	2340	2	B71704	cell surface anti
21	142.5	5.0	1072	2	A86827	hypothetical prote
22	142.5	5.0	2526	2	T20531	hypothetical prote
23	142.5	5.0	6658	2	T13931	endo-1,4-beta-xyla
24	142	4.9	1595	2	T31082	filamentous hemag
25	141.5	4.9	4152	2	T31102	hypothetical prote
26	141	4.9	1658	2	S55101	hypothetical prote
27	140.5	4.9	2539	2	B71619	SIR4 protein - yea
28	139	4.8	1358	2	A29360	reticulocyte-bind
29	138.5	4.8	1252	2	B42771	

30	138.5	4.8	1928	2	S46773	myosin heavy chain
31	138	4.8	1033	2	T37715	actin-interacting
32	137	4.8	1946	2	JC6032	lactocoeptin (EC 3.4
33	136.5	4.8	805	2	G82884	hypothetical prote
34	136.5	4.8	821	2	S67087	hypothetical prote
35	136	4.7	1465	2	S31262	TYB protein - yeas
36	136	4.7	1790	2	S67593	transport protein
37	136	4.7	1803	2	S56894	TYB protein - yeas
38	135.5	4.7	3724	2	T18427	hypothetical prote
39	135	4.7	1802	2	S52611	TYB protein - yeas
40	134.5	4.7	624	2	PC6003	surface membrane p
41	134.5	4.7	1558	2	B71603	RESA-H3 antigen PR
42	134	4.7	753	2	S48267	probable membrane
43	134	4.7	1570	2	T18272	1-phosphatidylinos
44	134	4.7	1711	2	T18429	hypothetical prote
45	134	4.7	2722	2	T20532	hypothetical prote

ALIGNMENTS

RESULT 1
I39934
Protective antigen precursor - Bacillus anthracis plasmid
C:Species: Bacillus anthracis
C:Date: 19-Jul-1996 #sequence, revision 19-Jul-1996 #text_change 01-Dec-2000
C:Accession: I39934; S69160; F59104
R:Weikos, S.L.; Lowe, J.R.; Eden-McCutchan, F.; Vodkin, M.; Leppla, S.H.; Schmidt, J.
Gene 69, 287-300, 1988
A:Title: Sequence and analysis of the DNA encoding protective antigen of Bacillus ant
A:Reference number: I39933; MWID:8912073
A:Accession: I39934
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-764 <RES>
A:Cross-references: GB:M22589; NID:g143280; PIDN:AAA22637.1; PID:g143282
R:Friedman, T.C.; Gordon, V.M.; Leppla, S.H.; Klimpel, K.R.; Birch, N.P.; Lob, Y.P.
Arch. Biochem. Biophys. 316, 5-13, 1995
A:Title: In vitro processing of anthrax toxin protective antigen by recombinant PC1 (
A:Reference number: S69160; MWID:95142670
A:Accession: S69160
A:Molecule type: protein
A:Residues: 197-202 <PRT>
R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koeh
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pXOI, the large Bacillus anthracis plasmid harb
A:Reference number: A59091; MWID:99445483
A:Accession: F59104
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-313, 'O', 315-764 <OKT>
A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32414.1; PID:g4894326
A:Experimental source: strain Sterne
A:Note: similar to anthrax toxin moiety, protective antigen, pagA formerly pag, plasm
C:Genetics:
A:Gene: pXOI-110
A:Genome: plasmid
C:Function:
A:Description: three component exotoxin; protective antigen binds to receptors on the
Y active components edema factor or lethal factor; the complex is internalized by rec
C:Keywords: exotoxin
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-196/Domain: propeptide #status predicted <PRO>
F:197-202/Product: protective antigen #status experimental <MAT>

Query Match 96.1%; Score 2758; DB 2; Length 764;
Best Local Similarity 96.8%; Pred. No. 4.7e-144;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;
Oy 1 DNQLPELKOKSSNRKRSTAGPTVPDRDNDGIPSLVEGYTVVKKRFLPSWIS 60
|||||
Db 178 DNQLPELKOKSSNRKRSTAGPTVPDRDNDGIPSLVEGYTVVKKRFLPSWIS 237

61 NHEKGLTKKSSPEKSTASDPYSDPEKTYGRIDKNVSEARHPPLVAAYPIVHVDEN 120
238 NHEKKGTLTKKSSPEKSTASDPYSDPEKTYGRIDKNVSEARHPPLVAAYPIVHVDEN 297
121 IILSNEQSONONTSEMTSTKNTSTRTHTSEVHGNAEVANSTRTHTSEVHGNAE 180
298 IILSNEQSONONTSEMTSTKNTSTRTHTSEVHGNAEVANSTRTHTSEVHGNAE 357
181 VHAVAIDHSLAGERTVAETMGLTADTARLANIRYVNTGTAPIVYVLPPTSLVLGN 240
358 SSTVAIDHSLAGERTVAETMGLTADTARLANIRYVNTGTAPIVYVLPPTSLVLGN 417
241 QTLATIKKENOQLOIAPNNYPSKNTAPALNAQDPSSTPTMANYNOELEKTKOL 300
418 QTLATIKKENOQLOIAPNNYPSKNTAPALNAQDPSSTPTMANYNOELEKTKOL 477
301 RLDPOVGNATYVFEAGRVAVDGSNMSEVLPQIOETATIIINGKDLNVERRIAY 360
478 RLDPOVGNATYVFEAGRVAVDGSNMSEVLPQIOETATIIINGKDLNVERRIAY 537
361 NPDPLETTKPDMLKEALKAIFGNEPENGLOYGKDTTEPFDFPDQOTSNIKOLAE 420
538 NPDPLETTKPDMLKEALKAIFGNEPENGLOYGKDTTEPFDFPDQOTSNIKOLAE 597
421 LNATNIVYVLKIKLNAMNILLRDKRFHYDNNIAGADESVYKHAHREVINSSTGL 480
598 LNATNIVYVLKIKLNAMNILLRDKRFHYDNNIAGADESVYKHAHREVINSSTGL 657
481 LNIDDKIRIISGYIVETEDREGKEVINDRYDMNIISSLRDGGTFIDFKYNDKPLX 540
658 LNIDDKIRIISGYIVETEDREGKEVINDRYDMNIISSLRDGGTFIDFKYNDKPLX 717
541 ISNPYKVNVAAYTKENT 558
718 ISNPYKVNVAAYTKENT 735

RESULT 2
140862
Iota toxin component Ib - Clostridium perfringens
C:Species: Clostridium perfringens
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Oct-1999
C:Accession: I40862; S42774
R:Perelle, S.; Gilbert, M.; Boquet, P.; Popoff, M.R.
I:Infect. Immun. 61, 5147-5156, 1993
A:Title: Characterization of Clostridium perfringens Iota-toxin genes and expression in
A:Reference number: I40861; MUID:94041637
A:Accession: I40862
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-875 <RES>
A:Cross-references: EMBL:X73562; NID:929031; PIDN:CAA51960.1; PID:9414655

Query Match 24.9%, Score 715.5; DB 2; Length 875;
Best Local Similarity 32.6%, Pred. No. 8.5e-32;
Matches 205; Conservative 102; Mismatches 207; Indels 115; Gaps 26;

1 DNLQJELKOKSSNRK-----KRSTAGPTVP-----D 29
160 DNLSPKLYLWELNGKTYIPENLFFRDYSKIDENDPFPNNFFDVAFFSAWEDDL 219
30 RDNDGIPDSELEGGYTVVKNKRTFLSPWISNHEKGLTKKSSPEKSTASDPYDPE 89
220 TDNDNIPAYEKNGTYI-----KDSIAVKMNSDFAE-QGYKKYVSSYLSNNTAGDPYTDYQ 274
90 KYTGRIKDNVSEARHPPLVAAYPIVHVDENIILSKNEQSTONTSETRTSTRTSTSR 149
275 KASGIDIAKLEADPLVAAYPVYGVEMENIISTNEHASS---DQKTVSRAFTNSK 330
150 THTSEVHGNAEVANSTRTHTSEVHGNAEVANSTRTHTSEVHGNAEVANSTRTHTSEVHGNAE 201

331 TDANTV--GVSTISAGYONGFTGNITTSYSHHTD-----NSTAVODS---NGE-SMWTG 377
202 MGLNTADTARLANIRYVNTGTAPIVYVLPPTSLVLGNKOTLATIKAKENOLQIAPNN 261
378 LSNKGSAYIANRYVNTGTAPIVYVLPPTSLVLGNKOTLATIKAKENOLQIAPNN 436
262 YPSKNTAPALNAQDPSSTPTMANYNOELEKTKOLRLDPOVGNATYVFEAGRV 321
437 YPSKNTAPALNAQDPSSTPTMANYNOELEKTKOLRLDPOVGNATYVFEAGRV 395
322 RVDGSMNSSEVLPQIOETATIIINGKDLNVERRIAYVNSDPLETTKPDMLKEALKI 381
436 -ITBGSNWSNYSQIDSVASATIID-7GSGQTFERRVAAKEDGCPNPKDT-PETTIGELAK 552
382 AFGNEPENGLOYGKDTTEPFDFPDQOTSNIKOLAEALNATNIVYVLKIKLNAMN 438
553 AFSATK-NGELLTFNGIFIDESCVELLIDDNSTSEIKKQKLYLDKTKIYV----KLERG 607
439 MNILIRDKRF--HYDR-NNIAGADESVYKHAHREVINSSTGL-----LNIDKD 486
608 MNILIRDKRF--HYDR-NNIAGADESVYKHAHREVINSSTGL-----LNIDKD 486
487 IRKILSGY-----IVEIEDREGKEVINDRYDMNIISSLRDGGTFIDFKYNDKPLX 521
666 KRYVSGYSKDPSTNSITVNIKSKEQKTDYLVPEKDYTKFSYEFETGKDSIDIEITLT 725
522 QDGKTFIDFKYND--KLPLYISNPYK 548
726 SSGVIFLNLSTELNSTPELKEPEIKV 754

RESULT 3
139933
Cryptic protein - Bacillus anthracis
C:Species: Bacillus anthracis
C:Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 15-Oct-1999
C:Accession: I39933
R:Welkos, S.L.; Lowe, J.R.; Eden-McCutchan, F.; Vodkin, M.; Leppla, S.H.; Schmidt, J.
I:Gene 69, 287-300, 1988
A:Title: Sequence and analysis of the DNA encoding protective antigen of Bacillus ant
A:Reference number: I39933; MUID:89172073
A:Accession: I39933
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-192 <RES>
A:Cross-references: GB:M2589; NID:9143280; PIDN:AAA22636.1; PID:9143281

Query Match 7.4%, Score 211.5; DB 2; Length 192;
Best Local Similarity 34.8%, Pred. No. 4.5e-05;
Matches 46; Conservative 31; Mismatches 34; Indels 21; Gaps 4;

439 MNILIRDKRFHYDRNNIAGADESVYKHAHREVINSSTGLNIDDKIRIISGYIYEI 498
1 MNILIRDKRFHYDRNNIAGADESVYKHAHREVINSSTGLNIDDKIRIISGYIYEI 59
499 EDTE-----GLKEVINDRYDMNIISSLRDGGTFIDFKYNDKPLIYSPN 545
60 KKPNSHLNLSPTITLACKDSVGLRYVLS-----DGAGTLDENKFDENMRSLY-DPG 112
546 YKVNVAAYTKEN 557
113 DDVYVVAAYTKED 124

RESULT 4
55104
Hypothetical protein pXOI-111 - Bacillus anthracis virulence plasmid pXOI
C:Species: Bacillus anthracis
C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 11-May-2000
C:Accession: G59104
R:Okumura, R.T.; Cloud, K.; Hampton, O.; Hofmaster, A.R.; Hill, K.K.; Koehn
I:Bacteriol. 181, 6509-6515, 1999

Query Match	6.38;	Score 181.5;	DB 2;	Length 2178;
-------------	-------	--------------	-------	--------------

```
OY      120 NIISKN--EDOSTONTDSETRTISKNTSRTHTEVHCNAEVAHANTSTRHTSEVG 1778
          ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB      825 NIYLNNKRTGEGVNSDGGANITFKASDNITMDGLNYDAE-----TVTKMIQT 875F
OY      178 NAEVAVAIDHSL-----SLAGERW-----AETWGLATAD-----TAR 2111
```

Db 876 GASQSHYAFDALNINISVNSSSDMTWKGFSFAKNISFNSRGCTNPGSSVISAN 935
Qy 212 LMANIRYVNT--GTAPIYVLPPTSLVLGKNQTLATI-----KAKENQLSQILAPANNYP 264
Db 936 ATNLSIFINSRLNGAVYN--LQANSLIFNNQAVFNVLSRGSNNATTOGLGNNFTL 994
Qy 265 S-----KLAIPAL--NAQDDSSPTITMANTNOELEKTYQLRLDQVYG 309
Db 995 SSQSLNFGNDTTLQNNANITLGNKSOAAFKNS--LTLDNNSNLSDNOSVLANNNNTSAFN 1053
Qy 310 NATYFENGGRVVDGSMSEVLPIQIETTARIIFNGKDLN--VERRTAAVNPSPLEET 368
Db 1054 NQASLNTIYGS-----QATENSLEFNGGTLSLMASSKLNASNASFSNNT 1097
Qy 369 TRPDMTLKALKIAGFNPNNGNLQYQKDIETEF-----DEFNDOOTSQINIKOLA 419
Db 1098 T---INLDDSVLASMTSSLNANINFGASQADFGNGTIIIDTASFPDASSLNFNNLLA 1154
Qy 420 -----ELNATYIVYLDKIKLNAKNNILIRDKRF- 448
Db 1155 NCALNFNGYTPSLTYALMSVSGQFVLGNNGDINLSDI--NIFDNITKSVTYNILLNAQKGIT 1213
Qy 449 -----HYDRNNTAVGA-----DESVYKFAHR-----EVIN- 473
Db 1214 GISGANGYEKILFYGMKIQNATYSDNNNTQTSFNPPLMSQIIOBSIKNGDLTIEVLNN 1273
Qy 474 -SSTEGLLINIDKI-----RKILSGYIYEIEDTEGLEKVINDRYDMLNSSL----- 520
Db 1274 PMSASNTIINIAPELYNQAOKNPFGYSYSDNQ-----GYTYLTSNIGLFPKGS 1328
Qy 521 --RODGKTFIDKKYNDKPLX---ISNPNYVNVYAVTKENT 558
Db 1329 QTPQAPGYTSPNPQLSSLNINYNKGFSSENLK--TLGLGISONS 1370

RESULT 7
P82885
hypothetical protein U0482 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: F82885
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
A:Reference number: A82870
A:Accession: F82885
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4688 <GLA>
A:Cross-references: GB:AE002145; GB:AF222894; NID:96899476; PIDN:AAF30894.1; GSPDB:GNO01
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: U0482
A:Genetic code: SGC3

Query Match 5.5%; Score 158; DB 2; Length 4688;
Best Local Similarity 21.2%; Pred. No. 4.1;
Matches 138; Conservative 107; Mismatches 270; Indels 136; Gaps 32;

Qy 1 DNLQJBLKOKSSN---SRKRSTSAGPTVDRD-----NDGIRD 37
Db 3803 DNLN--PETRYKLENIELSKPLKTHNLVSINDKENISLITETGNPVLKIOTONDITND 3861
Qy 38 SLEVEGYVDVKNKRTFLSPMISNIHEKGLTKYKSSPEKSTASDPYDFEKVYGRIDK 97
Db 3862 TQQTINVLISGVNKR--YNGRQIKVYKKNNNYTES--LITLQKGDYDOLLNLNS 3917
Qy 98 NVSPARHPPLVAAYPIVHV-----DME-----NIIISKNEOSTONTDSE-----TR 139
Db 3918 N-----REVREFEIKIINHISNTNPFDELEKLGVSNTFTTQTKNTTYVQNDSSATIVGR 3972

Qy 140 TISKNTSTSRHTSEVHGAEV---HANTSTSRHTSEVHGAENAVAHAIIDSLSLAGER 196
Db 3973 GVAFNFKI--KSEDKILENNQOVAVAPAKKETIHDNTWLOYPRLPKDVSDFK-----EG 4026
Qy 197 TAAETGGLT---ADPARLANIRYVNTGTAPIYNVLPPTSLVLGKNOLA-----TIK 247
Db 4027 TWAHDLSNVNFKETTYLVKIOFVNKPAAKNNINSENNVILDNSTINSNVEFTTK 4086
Qy 248 AKENQSLIAPANNYPKSNLAPIALN--AODESSPTIMNV--NOFLELEKTKOLRLPT 304
Db 4087 VGDHKLINTSSNNVNTFQNTIFTLISGVKKSVMGKKIKLSTKSDTSIHTNEVLIS 4146
Qy 305 DOVYGNIAFYENGR-----VRDTSGNMSEVLPIQIETTARIIFNGKDLNLEERRIA 358
Db 4147 NKTQYNILLNNLKRNTTYTLIDVKLIDNNNVSDFPEKGNLITNSFITRPSAINVLIEEI 4206
Qy 359 AVNPSPLEET-----KPDMTLKEA-----LKIAPFNENGMLQYQKDIETEF 402
Db 4207 SNRASNLSKSTIIKILNDPNDVLRDKQDATTYYGNNKQAMGFTYVSGNIKYLTLATVLDL 4266
Qy 403 DEFNDOOTSQNIK--NOLAEINATYITVLDKIKLNAKNNILIRDKRFHYD-----RNN 454
Db 4267 NFN--DKVINIVNISPNKKPSIAEN-----IGDKSNII-----YNNDSIPKLEIND 4313
Qy 455 IAVGA---DESVYKFAHREVINSSTEGLLINIDKIRKILSGYIYEIEDTEGLEKVIN- 510
Db 4314 IIVNGPINKIEIYVKNANOK--NNIDVDLGIQINPKIAHMLR--FIKFKSTN-----NDI 4364
Qy 511 -RYDMLNSSL--RODGKTFIDKKYNDKPLXISNPNYK--VNVYAVTKENT 558
Db 4365 IETNVINGSSLVNNDGKTSIRFTLNNLK-----ANKLYSLVDVYLVNNS 4410

RESULT 8
E71622
probable membrane associated protein PFB0125c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: E71622
R:Gardner, M.J.; Tellelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
; Pettes, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2, sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: E71622
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1308 <GAR>
A:Cross-references: GB:AE001374; GB:AE001362; NID:93845100; PIDN:AACT1815.1; PID:9384
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0125c

Query Match 5.5%; Score 157; DB 2; Length 1308;
Best Local Similarity 19.9%; Pred. No. 0.72;
Matches 130; Conservative 95; Mismatches 219; Indels 210; Gaps 31;

Qy 47 DYKNNRTFLSPMISNIHEKGLTKYKSSPEKSTASDPY----- 85
Db 42 DYKKDDVLT-----NIDKKKNVEEYKKNKQDVKNEDNEFFYFDKKEIKLKKIKEBOCN 96
Qy 86 ---SDFEKYTGRI--DKNVSPEARHPPLVAAYPIVHVDMENIILSKNEOSTONTDSETRT 140
Db 97 MKNEFINEKGYIILDENVS--TINNITSLNNDILHSSDNCVCTSYNIYPSNNNNNN-- 153
Qy 141 ISKNTSTSTRTHTSEVHGAENAVANTSTRTHTSEVHGAENAVAAIIDSLSLAGE----- 195
Db 154 -----NNNNVHISNNS-----NIFVDSHMEHDDITDEFFKID 187
Qy 196 RTWAETMGLNTADPARLANIRYVNTGTAPIYNVLPPTSLVLGKNQTLATIAKKNQISO 255
Db 188 QTNSEFPQFNISFENKKNVNEBELMKKHTDNT--NT---CDKIIDK-----KKNCNTLSD 236

```

Query Match          5.2%; Score 150.5; DB 2; Length 2269;
Best+Local Similarity 20.3%; Pred. No.3.6;
Matches 149; Conservative 87; Mismatches 250; Indels 249; Gaps 35;

OY 1 DNLQPELKKOKSSNRKRKSTASGTPVD-----RDNDGIPDS-LEVEGYTV-----46
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 204 DKTKLENLRSKIDNYDRIQKMEIETVASHLNIETNKKLPPTLIEIKKIYIDEISKELN 263
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 47 ----DVKKRKRFLEFWISINIEHK-GLTKIKSSPEKKSATSDPYSDFEKVTGRIDKNVSP 101
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 264 KMLEDFKKREKELSKIDYDKKREQLSEYKS--KMLEIRNHVYS-----OTVNDTKEE 316
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 102 EARHDLVAALPIVHDMEN---IILSKNDOSTONTDSETRTISKSTSRHTSEVHG- 157
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 317 EAKQ-----NDKSKNEHNTITPTNDE-----ISKILSEVTKMRKEILISK 356
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 158 -NAEYHANTSTSRHTSEVHG-----NAEVHVAVIDHSL-----SLAGERTWA- 199
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

	Query Match	5.28;	Score 150.5;	DB 2;	Length 2399;	
	Best Local Similarity	18.8%;	Pred. No. 4;			
	Matches 111;	Conservative 84;	Mismatches 201;	Indels 193;	Gaps 25;	
QY	120 NIIISKN--EEOGSRQNDDSERTRTISKNTSRTSKTTSEVHGNAEVHANTSTRHTHSVHG	177				
	: :	: :	:	:		
Dd	696 NIYLNNKTEGEGSNSGGGANITFRASDITMDGLNYYNAE-----TVTKLIQT	746				
QY	178 NAEVHANA---IDHSLSLA-----CERTW-----AETMGUN	205				
	: :	: :	: :	: :		
Dd	747 GASOHSYTTFEDATNNISVTSDDFSDMTWGFSPSAKANISFSNASFSGFTNPGGSSITSTN	806				

```

QY 206 TADT-----ARLNANIRY-----VNTGTAPIYVNL-----PTTSVLGKNQTLA 244
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 807 ASNSLSFTDSRLNGALINLQANSILFNNTQAVFNLSRGTSNPNATQOLLGNSTFLS 866
QY 245 TTKAKENQLSQILAPNNYPSKNIAPIAL--NAODEFSPTITMNYNOFLELEKTKQLRL 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 867 S-----QSLNFRNGDTTLQNNANITLGNKSOAFKNS--LTLLNNSKLSLDNOSVLNA 917
QY 303 DTDQVYGNATATYFNPNNGRVRVDTGSNWSEVLPQIOETTARIIFNGKDLNVERRIAANVP 362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 948 NQTSAPFNNOQASLINYNS-----QAFFSLFFNGTSL--NANSKILNA 959
QY 363 SDPLETTTPDMTLKALKIAGFENEPNGNLOYOQKDIET-----DENFDQOTSON 413
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 960 SSASFSPNMTTILMDSDVLANANTSSLNANINIFOGASQADFGGNTTIDYASFNFDSASSLN 1019
QY 414 IKNOIA-----ELNATITIVYLDKIKIMAKMNILL 443
Db 1020 FNNLTANGALNENGYAPSLTKALMNVSGQFVLGNNGDINLSDI--NIFDNTKSVTYNILL 1078
QY 444 RDKRF-----HYDRNRIAVGA-----DESVKFAHR----- 469
Db 1079 AQKGTIGISGANGYEKILFYGMKIQMATYSQNNNITQWTFINPLNSQIQIESIKNGDLT 1138
QY 470 -EVIN--SSTBGLIND-----KDIRKILSGYVEIEDTEGLKEVINDRYDMNLISSL 520
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1139 IEVLNPNASANTPINIAPELYNODSKONPTGSGYSDNQA-----GTYLYLTSTNIKGL 1193
QY 521 -----RODGTFIDFKKYNDRKPLV---ISNPNKYVNAVYAKENT 558
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1194 FTPKSGQTPQTPGTISPNQPLNSLNTINIKGFSSENLK--TLGLITLSDNS 1241

```

```

RESULT 11
T28317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T28317
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: Z20464; MUID:99102612
A:Accession: T28317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1127 <AF05>
A:Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97677.1; PID:g4049717
C:Genetics:
A:Note: MSV156

```

```

Query Match 5.2%; Score 150; DB 2; Length 1127;
Best Local Similarity 20.8%; Pred. No. 1.4; Mismatches 226; Indels 114; Gaps 28;
Matches 116; Conservative 103;
QY 44 YTVDKRKRTPSLPWSINHEKKGTLTKYSSPEKSTASDPYDEPK-VTGR---ID--- 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 YDLDEEKKD--KELVINIEKNAVDKINDIK--NNNNIHSMDNETIITGETLIDILN 171
QY 97 ---KNVSEARHPVAAVPIYH---VDMENI-----ILSKNDOSTONTSETPTISKNT 145
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 KLLKLVSSDEKQLLEQIKYNNKKEIFKNIDNYQKEINKKODELKLDESKKFEIKK- 230
QY 146 STSRHTSEVGNAEVHANTSTSRHTSEVGNAEVHNAVIDHSIAGERTWATMGLN 205
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 221 -----OELNKTIDKQDELIKRLNDEKINENIDEKOKLDOINS-----KIN 273
QY 206 TADTARLANIRYV--NTGTAPIYVNLPTTSVLGKNQTLATIKAKENQLSQILAPNNYYP 264
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 274 T-----LNENIKGVANLYTETKRNKISNQNELINKDSTIISKDEKOKLDELDKNINNT 328

```

```

QY 265 S-----KNLAPIALNADDEFSPTITMNYNOFLELEKTKQLRLDTP-OVYGNATY 314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 329 SLVYKSNMRTKITPQIOLLESSLIDFNNANININ-----EL-KSKIKLFPNDIOKLNNDDTE 382
QY 315 NFNNGRVRVDTGSNWSVLPQIOETTARIIFNGKDLN--VERRIAANVSPDLETTYPD 372
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 --QNNKIT--TDFPNNSTRIPEKRLDEYKTKIDIKNNNNLOKLEESYKKIDQOTEYKKNIN 439
QY 373 MTLKALKIAGFENEPNGNLOYOQKDIETEFDENFDQOTS-----QNTKNOLAELNATNI 426
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 440 KEYNDIIEI-----KNNNLQ-----KLEENKIDQOTEYKKNKINKEYNDIIEKNNNL 489
QY 427 YTVLDKIR-LNAKNMILIRDKRFH--YDRNRIAGADESVYKKAHREVINSSTE----- 477
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 490 OKLEENKININDIKLTKLKNIDIESNTELEFNKLN-----SDPKDKRSRIAKLNTIEYQLR 543
QY 478 -GLLINDIK--DIRKILSGYVEIEDTEGLKEVINDRYDMNLISLRDQKTFIDFKKYN 534
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 544 KDLLENINKTNEMLKLSDNKLSLSLEQLYDSKKNILDGID-----KIYNSLKEKN 592
QY 535 DKLPIYSNPYKVVNAVY 553
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 593 DKIDEIFSNIE-KFDIYV 610

```

```

RESULT 12
S56850
SMC1 protein homolog YJL074c - Yeast (Saccharomyces cerevisiae)
N:Alternate names: probable membrane protein YJL074c; protein J1049
C:Species: Saccharomyces cerevisiae
C:Date: 05-May-1995 #sequence_revision 08-Sep-1995 #text_change 29-Sep-1999
C:Accession: S56850; S57737
R:Rose, M.; Koeltter, P.; Entlian, K.D.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56848
A:Accession: S56850
A:Molecule type: DNA
A:Residues: 1-1230 <ROS>
A:Cross-references: EMBL:Z49349; NID:g1008230; PIDN:CAA89366.1; PID:g1008231; MIPS:YJ
R:Sor, F.J.
submitted to the EMBL Data Library, June 1995
A:Reference number: S57731
A:Accession: S57737
A:Molecule type: DNA
A:Residues: 1-1230 <SOR>
A:Cross-references: EMBL:X88851; NID:g895892; PIDN:CAA61313.1; PID:g895899
C:Genetics:
A:Gene: SGD:SMC3
A:Cross-references: SGD:S0003610; MIPS:YJL074c
A:Map position: 10L
C:Superfamily: hypothetical protein YJL074c
C:Keywords: transmembrane protein

```

```

Query Match 5.2%; Score 150; DB 2; Length 1230;
Best Local Similarity 20.4%; Pred. No. 1.6;
Matches 138; Conservative 102; Mismatches 259; Indels 182; Gaps 28;
QY 3 LQLEPKOKSSSRKKRSTASGPTVPDRDNDGIPDSLVEGTYVYKRRKFLSPWISNI 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 371 LQLASLQOKORLILIKKEIVAFKSKDRDPTWHSIEI-----ELKSS-----IONL 417
QY 63 HE-KKGLKRYKSSPEKWTASDPYSDPEKVTGRIDKNVSPEARHPVAAVPIVHVDENI 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 418 NELESQLOMDRTSLKQYSAID-----EIEIEILIDISNGPTQKGL-----EDPDSLI 466
QY 122 ILSKNEDOSTONTSETPTISKNTSTSRHTSEVGNAEVHANTSTSRHTSEVGNAEV 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 467 HLKQKLSLSDTRKELMRKEQKLOVLETLISDVNON--QRNVNETMSRS-----LANGII 520
QY 182 HAVADHSLAGF-----RTWATMG-----LWATDARLANAIR 217
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 521 NVKEITERLKISPESVFTGLGELLKRVNDKYRTCAEAVIGNSLFIHIVDTEATLIMNEL 580

```


[illegible]

RESULT 13
T31105
hypothetical protein 2 - Haemophilus ducreyi
C:Species: Haemophilus ducreyi
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31105
R:Ward, C.K.; Lumbley, S.R.; Letimer, J.L.; Cope, L.D.; Hansen, E.J.
J. Bacteriol. 180, 6013-6022, 1998
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein
A:Reference number: 420984; MUID:99030326
A:Accession: T31105
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4919 <NAR>
A:Cross-references: EMBL:AF057656; NID:g3929021; PID:g3929023; PIDN:AAC79761.1
C:Genetics:
C:Gene: 19PA2

Query Match	5.2%;	Score 150;	DB 2;	Length 4919;
Best Local Similarity	21.4%;	Pred. NO. 12;		
Matches 135;	Conservative 93;	Mismatches 244;	Indels 158;	Gaps 30

[illegible]

```

Oy 227 YNVLETTSLVIGKNOTLATIFAKEKQUSQILAPNNVYSKMLAPALNAQODFSSTPIITM 286
Db 614 KKV-----KNVTL-----NNDSE-LAANN-----LSLVA-----SHNVTL 642
Oy 287 NYNOFLELEKT--KOLRL--DTDOVYGNIAITYFENGSRVBDTGSNMSEVLPJOIETT 341
Db 643 NNSKLSAQKADIKAVNLTLDNTE---LRAKMLDINSTITITNGTIAIGFANI--TTE 696
Oy 342 RIENGKPLNIVERRI-AAVNPSDPLETTKPKDPMILKEALKTAFGHNEPNGLYOQKODIT 400
Db 697 KLNNEKALILIAEONLNFETVNGSH--YENKGDIVSKDKATVTFESKS----- 741
Oy 401 EEDFNEQDOTSQONIKNOJLAELNATNI-----TYVLDIKIKLNK-----MNI 441
Db 742 --DFFSNOSKLVNNAOMOL-KVANNNFETISQGDITJLIGNVTLNMSGFTTNSGMLTIVKTL 798
Oy 442 LIRDRFHYDRNNIYAVGAD---ESVVK-----EAHREVINS----- 474
Db 799 DVGDIQNTNNGKGNLVGSDGLHKSTKLTINDQKLISTKLNLMISSADPIINNGTLGIEAL 858
Oy 475 --STEGLLNLNDKDIRKLTLSGIYEIEDTEBGLKEYINDRYM--LNISS---LRODGK 525
Db 859 KIATGNGTNNKEKAL--LAASNLDDISVAEKKKTFNNGTIESGKULNITNTGAEFLWDNA 916
Oy 526 TEIDFKRYNDKLPLEYISN-----PKYKVVN 550
Db 917 TIRSGVNLITITSGVNSNNGTLISNERLNI 946

```

```

RESULT 14
568218
botulinum neurotoxin type A, nontoxic/nonhemagglutinating - Clostridium botulinum (str
C:Species: Clostridium botulinum
C:Variety: strain NIH
C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 15-Oct-1999
C:Accession: S68218; S74301
R:Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A:Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin compone
A:Reference number: S67968; MUID:96096783
A:Accession: S68218
A:Molecule type: DNA
A:Residues: 1-1193 <FUJ>
A:Cross-references: EMBL:D67030; NRD:g2160224; PIDN:BA11050.1; PID:d1011710; PID:g11
A:Experimental source: strain NIH
A:Accession: S74301
A:Molecule type: protein
A:Residues: 1-13;145-155 <FUJ>
A:Experimental source: strain NIH
C:Genetics:
A:Gene: ant
C:Keywords: neurotoxin
F11-1193/Product: botulinum neurotoxin type A, nontoxic/nonhemagglutinating #status e

```

Query Match 5.2%; Score 149; DB 2; Length 1193;
Best Local Similarity 19.3%; Pred. No. 1.7;
Matches 130; Conservative 107; Mismatches 238; Indels 198; Gaps 34;

[illegible]

Db 620 KKENISMPRIIESYEIPNDMLGLPLNDLNEKLFNITSKNTATPKKIYYNFDOMWTQYYSQ 679
QY 233 --TSLVLKNOQLA--TIKAKENQLSOILAPNNYPSKNIAPALNA---QDDFSSTP 283
Db 680 YFDLICMAKRSLVLAQETLIKRIIOKKLSYLIGNSNISSDNLALMLTTNTLRDISNSQ 739
QY 284 ITMN-----YNOFLEKTKQLRLTDQYVG--NATYFENGRRV 323
Db 740 IAMNVDSPFLNNAICVESNIYKFIISF-----MEOCINNINIKTKEF---IOK 786
QY 324 DTGSNMSEVLPOIOETTAIIIFNGKDLNVERRIAANVSPDLFTTKPDMTKEAL---- 379
Db 787 CTNINEDERKQLINON---VFNSLDPEFLN-----IQMKSLPSEFTRALLIKETWYRE 837
QY 380 KIAFGNEPNNGN-----IQYQKDI--TEPDFNFD-----QQTSONIKQLAEL 421
Db 838 LVLVAFKPEGNVIGDASGKNTSIEY--SKDIGLYGINSDALYNGSNQISFSNDPEFN 896
QY 422 NATNIYTV-----LDKIKLANKM-----NIIIR 444
Db 897 GITSFSTYFWLRNLGKDTIKSKLIGSKEDNCWEIYFQDTGLVNMIDSNCKENIYLS 956
QY 445 D---KRFHY-----DR--NNIAGADESVY--KEAHREVINSSTEGLLINIDKIKIISG 493
Db 957 DVSNNSMWHYITISVDRLKEQLIFIDNULVANESIKELIINIYSSNISLSNNPSYIEG 1016
QY 494 YIVEIEDTEGLEVINDRYDMLNISSLRQDKTFIDFK-----YNDKLPYISNPN 545
Db 1017 LILINKPPTS--QEVLSNFEVLNNSYIRDSNERLEYNNTQYLVYVPSDKRICVKKQNN 1075
QY 546 YKVNYYAATKENT 558
Db 1076 --NIY-LTINNT 1084

RESULT 15

D86818
hypothetical protein ypil [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
C:Accession: D86818
R:Bohotin, A.; Wincker, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich
Genome Res. In press, 2001
A:Title: The complete genome sequence of the lactic acid bacterium.
A:Reference number: A86625
A:Accession: D86818
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-810 <STO>
A:Cross-references: GB:AE005176; NID:g12724549; PIDN:AAK05646.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: ypil

Query Match 5.2%; Score 148; DB 2; Length 810;
Best Local Similarity 21.1%; Pred. No. 1.1;
Matches 115; Conservative 80; Mismatches 203; Indels 146; Gaps 25;

QY 107 LVAAAPPI-VHVMENIILSKNEDOSTON-----TDSEFTISKMT 145
Db 28 LVTSLPLAVKADENSAVPPNPOTEAIVNPVNSSLAASDSTEGNTGSSQAEFRKREANS 87
QY 146 STSRHTSEVHGNAEVHANTSTSRHTSEVHGNAEVHAVAIDHSLSLAGERTAAETMGLN 205
Db 88 LTSQASTLL---SASSVSSSTSSAQTSSV--EASSNSSEVSASLSKASTKAPSVLP- 141
QY 206 TADTARLANAIRYVNTGTAPIYVNLPTTSLVL---GKNQTLATIKAKENQLSOILAP-- 259
Db 142 --DSSKANVTYIASSLSAGSTIINPALTNGSIQANGOPVISTEATGVSANDITGPAT 199
QY 260 -----NNYPSKNIAPALNAODFSSPTITMNTNQLELEKTKQLRLDQYVG 309

Db 200 LDGINITTKANNLVNNFLTPDGLHMSDNTQYIVY-----KQATG 240
QY 310 NIATYNE--ENGRAVVDY-----GSNMSEVLPOIOETTAIIIFNGKDLNVERRIAANVPS 363
Db 241 QISSNFGILGTGPTTLATSYTMNAGDSGRITNVGRTLA-----GTNLDLIYKIVISTDESS 295
QY 364 --DPLETTKPKMTLKEALK-----IAFGNEPNNGLOYQKDIITEPDFNF-- 406
Db 296 WQAPKESTS--DCPIGLAFTGEONIANSDGNSIVALYGANNNVNLNYQ---IVVHNTNFOI 351
QY 407 -----DQOTSONIKNOLAEIATNIYTVLDIKIKLANMNLIRDKRFRHDBRN--NI 455
Db 332 PYLASFTITTDIMAGVNTNLA-----NLTIVIK-----TTNLTATDSSIIYDTTIPDT 401
QY 456 AVGADESVYKKAHREVINSSTEGLLINIDKD-----IKRIISGYIVE 497
Db 402 DLNGQASLPYGGYLGV-----GFLSNFDYDFYSPAPARSGDSYQSGGVHVDLFGSALQ 455
QY 498 IEDTEGLEVIN--DRYDMLNISSLRQDKTFIDFKKYDKLPYISN--PNYKVN--VYAVT 554
Db 456 AHLTQVVRDIYVLYNLYDADENNQLIVPOHOFIWPENVSMMNP--VSNPHYQYQWQNOTT 513
QY 555 KENT 558
Db 514 KQNT 517

Search completed: December 2, 2001, 13:51:43
Job time: 338 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:48:42 ; Search time 112.1 Seconds
(without alignments)
368.715 Million cell updates/sec

Title: US-09-747-521-4_COPY_178_735

Perfect score: 2871
Sequence: 1 DNLQPELKOKSSNSRRKKRS.....LYISNPKNVYAVTKENT 558

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101:*

1: /SID8/gcgdata/geneseq/AA1980.DAT:*
2: /SID8/gcgdata/geneseq/AA1981.DAT:*
3: /SID8/gcgdata/geneseq/AA1982.DAT:*
4: /SID8/gcgdata/geneseq/AA1983.DAT:*
5: /SID8/gcgdata/geneseq/AA1984.DAT:*
6: /SID8/gcgdata/geneseq/AA1985.DAT:*
7: /SID8/gcgdata/geneseq/AA1986.DAT:*
8: /SID8/gcgdata/geneseq/AA1987.DAT:*
9: /SID8/gcgdata/geneseq/AA1988.DAT:*
10: /SID8/gcgdata/geneseq/AA1989.DAT:*
11: /SID8/gcgdata/geneseq/AA1990.DAT:*
12: /SID8/gcgdata/geneseq/AA1991.DAT:*
13: /SID8/gcgdata/geneseq/AA1992.DAT:*
14: /SID8/gcgdata/geneseq/AA1993.DAT:*
15: /SID8/gcgdata/geneseq/AA1994.DAT:*
16: /SID8/gcgdata/geneseq/AA1995.DAT:*
17: /SID8/gcgdata/geneseq/AA1996.DAT:*
18: /SID8/gcgdata/geneseq/AA1997.DAT:*
19: /SID8/gcgdata/geneseq/AA1998.DAT:*
20: /SID8/gcgdata/geneseq/AA1999.DAT:*
21: /SID8/gcgdata/geneseq/AA2000.DAT:*
22: /SID8/gcgdata/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2758	96.1	735	15	AA60179
2	2758	96.1	736	21	AA156959
3	2758	96.1	763	21	AA156960
4	2758	96.1	764	21	AA156960
5	2758	96.1	764	22	AA156958
6	2758	96.1	764	22	AA156958
7	2664	92.8	569	21	AA156961
8	2581	89.9	719	15	AA60193
9	593	20.7	884	17	AA60193
10	591.5	20.6	880	19	AA60224
11	591	20.6	881	21	AA159277

12	589	20.5	884	18	AA19509	B. cereus VIP1(a)
13	589	20.5	884	19	AA19512	100 kDa VIP1(a) t
14	589	20.5	1346	17	AA191245	VIP2(a) and VIP1A
15	589	20.5	1346	18	AA19513	B. cereus VIP1(a)
16	589	20.5	1346	19	AA19513	B. cereus VIP1(a)
17	588	20.5	852	17	AA191246	VIP1(a) protein w
18	588	20.5	852	18	AA19516	Maize optimised-B.
19	588	20.5	852	18	AA19516	Maize optimised-B.
20	588	20.5	1338	17	AA191247	VIP2(a)-VIP1(a)
21	588	20.5	1338	18	AA19520	VIP2(a)-VIP1(a)
22	588	20.5	1338	18	AA19520	VIP2(a)-VIP1(a)
23	584.5	20.4	884	15	AA19513	Bacillus cereus 10
24	584.5	20.3	667	17	AA191240	B. cereus VIP1 pro
25	583	20.3	860	21	AA195282	MIS-8 toxin from B
26	582.5	20.3	784	21	AA195282	MIS-8 toxin from B
27	580	20.2	667	18	AA19510	B. cereus 80 kD VI
28	580	20.2	667	19	AA19510	80 kDa VIP1(a) to
29	578	20.1	667	15	AA19513	Bacillus thuringie
30	572.5	19.9	834	17	AA191242	B. thuringiensis V
31	572.5	19.9	834	18	AA19512	B. thuringiensis V
32	572.5	19.9	834	18	AA19512	Vegetative insect
33	411	14.3	425	19	AA19512	Bacillus thuringie
34	411	14.3	425	19	AA19512	Bacillus thuringie
35	315	11.0	357	19	AA19512	Bacillus thuringie
36	307	10.7	357	19	AA19512	Bacillus thuringie
37	303	10.6	357	19	AA19512	Bacillus thuringie
38	291	10.1	357	19	AA19512	Bacillus thuringie
39	288	10.0	327	19	AA19512	Bacillus thuringie
40	278	9.7	327	19	AA19512	Bacillus thuringie
41	256.5	8.9	333	19	AA19512	Bacillus thuringie
42	252	8.8	347	21	AA19512	Bacillus thuringie
43	252	8.8	347	21	AA19512	Toxin from B. thur
44	252	8.8	348	19	AA19512	Bacillus thuringie
45	252	8.8	348	19	AA19512	Bacillus thuringie

ALIGNMENTS

RESULT 1	
AA19512	AA19512 standard; Protein: 735 AA.
XX	
AC	AA19512
XX	
DT	03-APR-1995 (first entry)
XX	
DE	Protective antigen of Bacillus anthracis.
XX	
KW	Anthrax; Bacillus anthracis; fusion protein; protective antigen;
KW	protective antigen; cell killing; targeting; pathogen;
KW	intracellular; HIV; human immunodeficiency virus; toxin.
XX	
OS	Bacillus anthracis.
XX	
PN	WO9418332-A.
XX	
PD	18-AUG-1994.
XX	
PF	14-FEB-1994; 94WO-US01624.
XX	
PR	12-FEB-1993; 93US-0021601.
PR	25-JUN-1993; 93US-0082849.
XX	
PA	(US) US DEPT HEALTH & HUMAN SERVICES.
XX	
PI	Arora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;
XX	
DR	WPI, 1994-279753/34.
DR	N-PSDB; AA070180.
XX	
PT	Nucleic acid encoding anthrax toxin fusion protein - useful for
	targeting toxin to specific cells, eg for killing tumour cells

or HIV-infected cells

Disclosure: Page 81-83; 124pp; English.

The sequence encoding the protective antigen of *Bacillus anthracis* may be used in the construction of a nucleic acid which encodes a fusion protein comprising the anthrax protective antigen binding domain of the native anthrax lethal factor and a sequence encoding an activity inducing domain of a second protein. The fusion proteins are useful for the specific killing of tumour cells or the killing of cells infected with intracellular pathogens, especially HIV.

Sequence 735 AA:

Query Match 96.1%; Score 2758; DB 15; Length 735;
Best Local Similarity 96.8%; Pred. No. 7,4e-188;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

```

QY 1 DNLQLEPKKQSSNRKRSTSGAPTPVDRNDGIPDSLEVEGYVVDVKNKRTPLSPWIS 60
   |||||||
DB 149 dnlqlpelkqksnsarkrstaagtpvdrndgipdslevegylvdknkrtflispwls 208
   |||||||

QY 61 NHEKKGLTKYKSSPEKKSTASDPYSDFEKYVGRIDKNVSPARRPLVAAYPIVHVDEN 120
   |||||||
DB 209 nhnekkgltkyasspekswtadpysdfekevgridknvspearplvaaypihvhdmen 268
   |||||||

QY 121 ILSKNEOSTONTSETRTISKNTSTKRTHTSEVHGNAEVANNTSTSRHTSEVHGNAE 180
   |||||||
DB 269 ilsknedsgntdsetrtiskntstkrthtsevhgnaevnastfdidgsvsagfsn 328
   |||||||

QY 181 VHAVAIDHSLSAGERTAEIWMGLTADTARLANIRIVNTGTAIVNVLPPTSVLGKN 240
   |||||||
DB 329 sstvaldhslsagertvaetmgltadtarlhanirvntgtaplrvlpptsivlgkn 388
   |||||||

QY 241 OTLATIRAKENQLSOILAPNNYPSKNLAPITALNODDESSPRIMANTNOFELEKTKOL 300
   |||||||
DB 389 qlatirakenkqlsqilapnnypsknlapitalnagddstscplmnyngfielektqjl 448
   |||||||

QY 301 RLDTQVYGNITATVNFENGVRVDTGSNMSEVLPOIQETTARIIFNGKDLNVERRIAAY 360
   |||||||
DB 449 rldtqvygniatyvfengvrvtgdsnmsevlpqigetarlifngkdlnverriaav 508
   |||||||

QY 361 NPSDPLETTKPDMTLKEALKIAFGFNEPNGNLOYGKDIITEFDENFDQOTSINIKNOAE 420
   |||||||
DB 509 npsdpletckpmtlkealkiafgfnepnngnlyqgskdtlfdnfdqtsqnknqlae 568
   |||||||

QY 421 LNATNITVVLDRKIKLNAKMNILIRDKRPHYDRNNTAVGADSEVKEAHEVINSSTEGIL 480
   |||||||
DB 569 lnatnityvlldrkiklnakmmllirdkrthydrrnmlavgadesevkeahrevinsstegil 628
   |||||||

QY 481 LNIIDKIRKILSGYIVETEDTEGLEKVINDRYDMLNISLRDQGTFTIDFKKYNKPLPLY 540
   |||||||
DB 629 lnidkirkilsgyiveledteglekevindr ydmlnisslrddgtftidfkknkplply 688
   |||||||

QY 541 ISNPYKYNVAVTKENT 558
   |||||||
DB 689 isnpykvyvavtkent 706
   |||||||

```

RESULT 2

AAVS6959 standard; protein: 736 AA.

AAVS6959;

25-APR-2000 (first entry)

B. anthracis MAT-PA protein.

Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA; tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.

Bacillus anthracis.

WO200002522-A2.

20-JAN-2000.

09-JUL-1999; 99WO-US15568.

10-JUL-1998; 98US-0092416.

(USME-) US MEDICAL RES INST INFECTIOUS DISEASES.

Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;

WPI: 2000-182165/16.

N-PSDB; AA256875.

Recombinant DNA construct useful as vaccines for anthrax, in producing host cells for analyzing the drugs and agents inhibiting anthrax -

Disclosure: Page 34; 35pp; English.

The invention provides a recombinant DNA construct that comprises a vector and at least one nucleic acid (or its fragment) encoding a combination of *Bacillus anthracis* proteins, selected from protective antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA (PA with its secretory signals replaced with those of tissue plasminogen activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine for anthrax and in producing infectious alpha virus particles. These particles, expressing the B. anthracis proteins are useful also as vaccines for anthrax. Host cells transformed with the construct are useful for analyzing the effectiveness of drugs and agents that inhibit anthrax or B. anthracis proteins. The present sequence represents a B. anthracis MAT-PA protein.

Sequence 736 AA:

Query Match 96.1%; Score 2758; DB 21; Length 736;
Best Local Similarity 96.8%; Pred. No. 7,4e-188;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

```

QY 1 DNLQLEPKKQSSNRKRSTSGAPTPVDRNDGIPDSLEVEGYVVDVKNKRTPLSPWIS 60
   |||||||
DB 150 dnlqlpelkqksnsarkrstaagtpvdrndgipdslevegylvdknkrtflispwls 209
   |||||||

QY 61 NHEKKGLTKYKSSPEKKSTASDPYSDFEKYVGRIDKNVSPARRPLVAAYPIVHVDEN 120
   |||||||
DB 210 nhnekkgltkyasspekswtadpysdfekevgridknvspearplvaaypihvhdmen 269
   |||||||

QY 121 ILSKNEOSTONTSETRTISKNTSTKRTHTSEVHGNAEVANNTSTSRHTSEVHGNAE 180
   |||||||
DB 270 ilsknedsgntdsetrtiskntstkrthtsevhgnaevnastfdidgsvsagfsn 329
   |||||||

QY 181 VHAVAIDHSLSAGERTAEIWMGLTADTARLANIRIVNTGTAIVNVLPPTSVLGKN 240
   |||||||
DB 330 sstvaldhslsagertvaetmgltadtarlhanirvntgtaplrvlpptsivlgkn 389
   |||||||

QY 241 OTLATIRAKENQLSOILAPNNYPSKNLAPITALNODDESSPRIMANTNOFELEKTKOL 300
   |||||||
DB 390 qlatirakenkqlsqilapnnypsknlapitalnagddstscplmnyngfielektqjl 449
   |||||||

QY 301 RLDTQVYGNITATVNFENGVRVDTGSNMSEVLPOIQETTARIIFNGKDLNVERRIAAY 360
   |||||||
DB 450 rldtqvygniatyvfengvrvtgdsnmsevlpqigetarlifngkdlnverriaav 509
   |||||||

QY 361 NPSDPLETTKPDMTLKEALKIAFGFNEPNGNLOYGKDIITEFDENFDQOTSINIKNOAE 420
   |||||||
DB 510 npsdpletckpmtlkealkiafgfnepnngnlyqgskdtlfdnfdqtsqnknqlae 569
   |||||||

QY 421 LNATNITVVLDRKIKLNAKMNILIRDKRPHYDRNNTAVGADSEVKEAHEVINSSTEGIL 480
   |||||||
DB 570 lnatnityvlldrkiklnakmmllirdkrthydrrnmlavgadesevkeahrevinsstegil 629
   |||||||

```

OY 481 LINDKDIRKILSGYIEIDTEGLKEVINDRYDMLNSSLRODGKTFIDFKKYNDKLPly 540
 |||||||
 Db 630 lndkdlrklilsgyiveidteglkeyindrydmnlsslrqdgkftidtkyndkrlply 689
 OY 541 ISNPNNKVVNVAVTKE NT 558
 |||||||
 Db 690 isnpnkvnvyavtkent 707

RESULT 3

AA56960
 ID AAY56960 standard; protein; 763 AA.

AC AAY56960;

DT 25-APR-2000 (first entry)

DE B. anthracis TPA-PA protein.

KW Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;
 tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.

OS Bacillus anthracis.

PN WO200002522-A2.

PD 20-JAN-2000.

PF 09-JUL-1999; 99WO-US15568.

PR 10-JUL-1998; 98US-0092416.

XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.

XX Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;

DR WPI: 2000-182165/16.

XX N-PSDB; AA256876.

PT Recombinant DNA construct useful as vaccines for anthrax, in producing
 host cells for analyzing the drugs and agents inhibiting anthrax -

XX Disclosure; Page 32; 35pp; English.

CC The invention provides a recombinant DNA construct that comprises a
 vector and at least one nucleic acid (or its fragment) encoding a
 combination of Bacillus anthracis proteins, selected from protective
 antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
 (PA with its secretory signals replaced with those of tissue plasminogen
 activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
 for anthrax and in producing infectious alpha virus particles. These
 particles, expressing the B. anthracis proteins are useful also as
 vaccines for anthrax. Host cells transformed with the construct are
 CC useful for analyzing the effectiveness of drugs and agents that inhibit
 anthrax or B. anthracis proteins. The present sequence represents a
 CC B. anthracis TPA-PA protein.

XX Sequence 763 AA;

Query Match 96.1%; Score 2758; DB 21: Length 763;

Best Local Similarity 96.8%; Pred. No. 7.8e-186;

Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

OY 1 DNLOJPELKOKSSNRKRSTSGAPTPVPRDNDGIPDSLEVEGYTVDVKNKRTFLSPWIS 60
 |||||||
 Db 177 dnlqlpelkqksnsrkrstsaagptvprdrndgipdsleveytydvknkrflfispwis 236
 OY 61 NIHEKGLTKRYSSPEKWSSTADSPYSDFEKKVYGRIDKNVSPEARHPLVAAYPIVHVDMEN 120
 |||||||
 Db 237 nlhekgltkrysspekwsstadsydfekvtgridknvspearhplvaayplvhwmden 296

OY 121 IILSKNEDOSTONWSETPTISKNTSTSRTHSEVHNAEVAHNTSTSRTHSEVHNAE 180
 |||||||
 Db 297 ilsknedgstqntdestrliskntstsrthsevhnaevhasffdgsvsaagfsn 356
 OY 181 VHAVALDHSLAGERTWAEFTMGINTADTARLANANIRYVNTGAPLYNVLPPTSLVGLKN 240
 |||||||
 Db 357 ssevaldhsislagerwaetmgjntadtarlanairyntgcaplynvlpptslvlgkn 416
 OY 241 QTLATIRAKENQSLQILAPNNYPSKNLAPIALNAODESSTPTMMYNQFLEEKTKOL 300
 |||||||
 Db 417 qlatlirakenqslqilapnnypsknlapialnaoddisstptlmynqflekckql 476
 OY 301 RLPTDOVYGNATYVNFENGVRVDTGSMNSEVLPQIETTARIIFNGKDLNVERRIAAY 360
 |||||||
 Db 477 rldtdgygniatyvnfengvrvdtgsmnsevlpqiettarilfngkdlnverriaav 536
 OY 361 NPSDPLETTRKDMTLKALKIAFGFNEPNGNLOYOGKDITEPFNFQDQTSQNIKNLOAE 420
 |||||||
 Db 537 npsdplettrkdmntlkealkiafgfnepngnlgygkditefdnfqgtsqnlknqlae 596
 OY 421 LNMTNITYVDKTKLNKMMILIRDKRFHYDRNNIANGADESVYKKAHREYINSGTGL 480
 |||||||
 Db 597 lnmtnityvdtklknkmmillrdrkrfhydrnniavgadesvvykkahevinsstegl 656
 OY 481 LINDKDIRKILSGYIEIDTEGLKEVINDRYDMLNSSLRODGKTFIDFKKYNDKLPly 540
 |||||||
 Db 657 lndkdlrklilsgyiveidteglkeyindrydmnlsslrqdgkftidtkyndkrlply 716
 OY 541 ISNPNNKVVNVAVTKE NT 558
 |||||||
 Db 717 isnpnkvnvyavtkent 734

RESULT 4

AA56958

ID AAY56958 standard; protein; 764 AA.

XX AAY56958;

DT 25-APR-2000 (first entry)

DE B. anthracis protective antigen (PA) protein.

KW Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;
 tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.

OS Bacillus anthracis.

PN WO200002522-A2.

PD 20-JAN-2000.

PF 09-JUL-1999; 99WO-US15568.

PR 10-JUL-1998; 98US-0092416.

XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.

XX Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;

DR WPI: 2000-182165/16.

XX N-PSDB; AA256874.

PT Recombinant DNA construct useful as vaccines for anthrax, in producing
 host cells for analyzing the drugs and agents inhibiting anthrax -

XX Disclosure; Page 33; 35pp; English.

CC The invention provides a recombinant DNA construct that comprises a
 vector and at least one nucleic acid (or its fragment) encoding a
 combination of Bacillus anthracis proteins, selected from protective
 antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
 (PA with its secretory signals replaced with those of tissue plasminogen

CC activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
 CC for anthrax and in producing infectious alpha virus particles. These
 CC particles, expressing the B. anthracis proteins are useful also as
 CC vaccines for anthrax. Host cells transformed with the construct are
 CC useful for analyzing the effectiveness of drugs and agents that inhibit
 CC anthrax or B. anthracis proteins. The present sequence represents a
 CC B. anthracis PA protein.

XX
 SQ Sequence 764 AA;

Query Match 96.1%; Score 2758; DB 21; Length 764;
 Best Local Similarity 96.8%; Pred. No. 7.8e-188;
 Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 DNLOPELKOKSSNRKRSTAGPTVPDRNDGIPDSLEVEGYVDVKNKRTFLSPWIS 60
 Db 178 dnldpelpkqgssnrkrstagsptvpdrndgipdslevegylvdvknkrtflispwis 237
 QY 61 NIHEKKGITKKYKSSPEKSTASDPYSDEKYTGRIKKNVSPGARHPLVAAYPIVHVDMEN 120
 Db 238 nihekkgltkykspekstasdpysdefkytgridknvspearhplvaaypivhvdmen 297
 QY 121 IILSKNEQSTQNDSETRTISKNTSTRTHTSEVHGNAEYHANTSTRTHTSEVHGNAE 180
 Db 298 ilsknedqstqndsetrtiskntsttrhtsevhgnaevhaasfidigsvsagfnsn 357
 QY 181 VHVAVIDHSLSLAGERTVAETMGLNTADTARLNANIRVYNTGTAPIYVAVLPTTSVLGKN 240
 Db 358 ssvaidslslagertvaetmglntadtarlannirvyngtapiyvavlpitsvlgkn 417
 QY 241 QTLATIRAKENOLSOILAPNNYPSKNLAPIALNAQDFFSSTPITMANYNOFLELEKTQOL 300
 Db 418 qtlatikakenglsqilapnnypsknlapialnaqddfsstptlmynqflelektqql 477
 QY 301 RLDDQVYGNATYVFNFGRVVDGSMWSEVLPQIOETTAIRIIFNGKDLNVERRIAAY 360
 Db 478 rlddqvygnatynfengrvvrdgsmwsevlpqioetarlirifngkdlnverriaav 537
 QY 361 NPSDPLETTKPDMTLEALKTAFGNEPNNGNLOYOGKDITFEDEFNDQSTQNIKQOLAE 420
 Db 538 npsdplettkpdmtlealktafgnepnngnloyogkditefdefndqstqniknqlae 597
 QY 421 LNAATNIVYLDKIKLNAMNLLIRDKRFHYDRNNINAVGADSVYKEAHREYINSTBGLL 480
 Db 598 lnatniyylvldkiklnamnlirdkrfhydrnniavgadesvvykeahreynsstegl 657
 QY 481 LNIKDRIKIIISGYIVETEDGELKEVINDRYDMNIISSLRQDGKTFIDFKKYNKCLPLY 540
 Db 658 lnikdrikiiisgyiveledgeglevindydmniiislrqdgkclidfkknkclply 717
 QY 541 ISNPNYKVNVAAYVAKENT 558
 Db 718 isnpnykvnvayvackent 735

RESULT 5
 AAB47306 ID AAB47306 standard; Protein: 764 AA.
 XX
 AC AAB47306;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE Wild type B. anthracis protective antigen.
 XX
 KW Lethal factor; LF; immunogen; LF4; protective antigen; PA; DNA vaccine;
 KW humoral; cell-mediated; immune memory response.
 XX
 OS Bacillus anthracis.
 XX
 PH Key Location/Qualifiers
 FT Peptide 1..29

FT /label= Signal peptide
 FT /note= "Not given in the specification"
 FT Protein 30..764
 FT /label= PA
 FT Peptide 204..764
 FT /label= pcpa
 XX
 PN W0200145639-A2.
 XX
 PD 28-JUN-2001.
 XX
 PF 21-DEC-2000; 2000WO-US34912.
 XX
 PR 22-DEC-1999; 99US-0171459.

PA (OHIS) UNITV OHIO STATE RES FOUND.

PA (GALL/) GALLOMAY D R.

PA (MATECZUN A J.

PI Galloway DR, Mateczun AJ;

DR WPI: 2001-408540/43.

DR N-PSDB: AAC86016.

PT Protecting animal against lethal infection with Bacillus anthracis, by
 PT administering wild type or mutated form of Bacillus anthracis lethal
 PT factor protein or its fragment or a nucleic acid encoding the mutated
 PT protein -

PS Claim 5; Fig 2; 33pp; English.

XX This sequence shows the B. anthracis protective antigen (PA). An
 CC immunogenic fragment of PA, pcpa, can be used to produce an immune
 CC response which protects an animal against lethal infection with
 CC Bacillus anthracis. DNA encoding the B. anthracis PA can be used
 CC in conjunction with DNA encoding the lethal factor (LF) in a
 CC DNA vaccine. Using a DNA vaccine which encodes the mutated LF protein
 CC or its fragment alone or in combination with a DNA encoding the PA protein
 CC or its fragment, both components (humoral and cell-mediated) of the
 CC immune system are stimulated, which results in longer term immune
 CC memory response. The combined use of a mutated LF and PA gene or their
 CC fragments results in a higher level of immune response, as judged by
 CC overall serum antibody titers for LF and PA antigens, than the use of
 CC either LF or PA genes in separate immunizations.

XX Sequence 764 AA;

Query Match 96.1%; Score 2758; DB 22; Length 764;
 Best Local Similarity 96.8%; Pred. No. 7.8e-188;
 Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 DNLOPELKOKSSNRKRSTAGPTVPDRNDGIPDSLEVEGYVDVKNKRTFLSPWIS 60
 Db 178 dnldpelpkqgssnrkrstagsptvpdrndgipdslevegylvdvknkrtflispwis 237
 QY 61 NIHEKKGITKKYKSSPEKSTASDPYSDEKYTGRIKKNVSPGARHPLVAAYPIVHVDMEN 120
 Db 238 nihekkgltkykspekstasdpysdefkytgridknvspearhplvaaypivhvdmen 297
 QY 121 IILSKNEQSTQNDSETRTISKNTSTRTHTSEVHGNAEYHANTSTRTHTSEVHGNAE 180
 Db 298 ilsknedqstqndsetrtiskntsttrhtsevhgnaevhaasfidigsvsagfnsn 357
 QY 181 VHVAVIDHSLSLAGERTVAETMGLNTADTARLNANIRVYNTGTAPIYVAVLPTTSVLGKN 240
 Db 358 ssvaidslslagertvaetmglntadtarlannirvyngtapiyvavlpitsvlgkn 417
 QY 241 QTLATIRAKENOLSOILAPNNYPSKNLAPIALNAQDFFSSTPITMANYNOFLELEKTQOL 300
 Db 418 qtlatikakenglsqilapnnypsknlapialnaqddfsstptlmynqflelektqql 477
 QY 301 RLDDQVYGNATYVFNFGRVVDGSMWSEVLPQIOETTAIRIIFNGKDLNVERRIAAY 360

```

Db      478 rldtdqygnlactyngfengrvrdtgsnwsevlpqiqettarilifngkdliverriaav 537
Oy      361 NPSDPLETTRKPMWTLKEALIKAFGENPENGNOYOGKDIRTEPFENDOOTSONIKNOLAE 420
Db      538 mpsdpletckpmtlkealkiafgrnepnqlyqgkdteidfndqgtsqnlknlae 597
Oy      421 LNATNIYTVLDKIKLNAKNNILIRDKRFHYDRNNINAVGADESYYKEAAREVINSSTEGLL 480
Db      598 lnatniyvtldkiklnaknnmlllrdrfhydrnnlavgadesvvekeahrevinsstegll 657
Oy      481 LNIDKDIRKILSGYIYEIEDTSGLKEVINDRIDMLNISSLRDQKTFIDPKKYNKLP 540
Db      658 lndkdrkrlisgyiveiedteglkevindrmdlnisslrqgkctfidfkynkldply 717
Oy      541 ISNPYKVNYYAVTKENT 558
Db      718 isnpykvnnyavtkent 735

```

RESULT 6

AAR60183 standard: Protein; 903 AA.

```

ID      AAR60183 standard: Protein; 903 AA.
XX
AC      AAR60183;
XX
DT      04-APR-1995 (first entry)
XX
DE      PA(1-725)-----Human CD4 fusion protein coding sequence.
XX
KW      Anthrax; Bacillus anthracis; fusion protein; lethal factor;
KM      protective antigen; cell killing; targeting; targeting; pathogen;
KW      intracellular; HIV; human immunodeficiency virus; coxin;
XX      Pseudomonas; exotoxin.
XX
OS      Bacillus anthracis.
XX
PN      Homo sapiens.
XX
PD      WO9418332-A.
XX
PF      14-FEB-1994; 94WO-US01624.
XX
PR      12-FEB-1993; 93US-0021601.
XX      25-JUN-1993; 93US-0082849.
XX
PA      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI      Aroza N, Klimpel K, Leppla SH, Nichols PJ, Singh Y.
XX
DR      WPI: 1994-279753/34.
XX      N-PSDB; AAQ70184.
XX
PT      Nucleic acid encoding anthrax toxin fusion protein - useful for
XX      targeting toxin to specific cells, eg for killing tumour cells
XX      or HIV-infected cells
XX
PS      Disclosure; Page 100-103; 124pp; English.
XX
CC      This sequence is a fusion protein comprising amino acid residues
XX      1-725 of the anthrax protective antigen protein and residues 1-178
XX      of human CD4, the portion which binds to gp120 on HIV infected
XX      cells. Such fusion proteins may be useful for the specific killing
XX      of tumour cells or the killing of cells infected with intracellular
XX      pathogens, especially HIV, depending on their components.
XX
SO      Sequence 903 AA.

```

Query Match 96.1%; Score 2758; DB 15; Length 903;
 Best Local Similarity 96.8%; Pred. No. 1e-187;
 Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

```

Oy      1 DNLQPELKOKSSNSRKKRSTASGPTVPDRDNDGIDPSLEVEGYVDVKNKRTFLSPWIS 60
Db      149 dnqlpelkksnsnrkkrtstasgptvpdrdndgipdslevegvydvknkrtflspwis 208
Oy      61 NIHEKGLTKYKSSPEKWSASPYSDFEYKTVGRIDKNVSEAKRHPLVAAPYIVHVDMEN 120
Db      209 nlhekgltkykspekwsaspydsdfekvtgridknvseaprhplvaapyivhvdmen 268
Oy      121 IILSKNEDOSTOMTDSFTRISKNTSRRHTSEVGNAAVHANTSRHRTSEVGNAAE 180
Db      269 iilsknedgstomtdsftriskntsrhrtsevgnaevhasfidlgsvsagfsn 328
Oy      181 VHAVALDHSLSLGEPTWATMGLNTPARLANIRYVNTGTAPYVNLPTTSLVLGKN 240
Db      329 ssvvaldhslsagertwætmglnctatrlanirvntgtcapynvlpptsylvlgkn 388
Oy      241 QTLATIRAKENQLSQILAPNNYPSKNLADIALNADDESSTPIYNNQFLELEKTKOL 300
Db      389 qtlatirkakenqslsapnnypsknladialnaqddestpilmynqflelektkql 448
Oy      301 RLDTDYVYGNIAIYVNFENGVRVDTGSNMSEVLPQIDETARIIFNCKDLNIVERRIAAY 360
Db      449 rldtdqygnlactyngfengrvrdtgsnwsevlpqiqettarilifngkdliverriaav 508
Oy      361 NPSDPLETTRKPMWTLKEALIKAFGENPENGNOYOGKDIRTEPFENDOOTSONIKNOLAE 420
Db      509 mpsdpletckpmtlkealkiafgrnepnqlyqgkdteidfndqgtsqnlknlae 568
Oy      421 LNATNIYTVLDKIKLNAKNNILIRDKRFHYDRNNINAVGADESYYKEAAREVINSSTEGLL 480
Db      569 lnatniyvtldkiklnaknnmlllrdrfhydrnnlavgadesvvekeahrevinsstegll 628
Oy      481 LNIDKDIRKILSGYIYEIEDTSGLKEVINDRIDMLNISSLRDQKTFIDPKKYNKLP 540
Db      629 lndkdrkrlisgyiveiedteglkevindrmdlnisslrqgkctfidfkynkldply 688
Oy      541 ISNPYKVNYYAVTKENT 558
Db      689 isnpykvnnyavtkent 706

```

RESULT 7

AAV56961 standard: protein; 569 AA.

```

ID      AAV56961 standard: protein; 569 AA.
XX
AC      AAV56961;
XX
DT      25-APR-2000 (first entry)
XX
DE      B. anthracis PA63 protein.
XX
KW      Bacillus anthracis protein; protective antigen; PA; MAT-PA; rPA-PA;
KM      tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.
XX
OS      Bacillus anthracis.
XX
PN      WO200002522-A2.
XX
PD      20-JAN-2000.
XX
PF      09-JUL-1999; 99WO-US15568.
XX
PR      10-JUL-1998; 98US-0092416.
XX
PA      (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI      Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL.
XX
DR      WPI: 2000-182165/16.
XX      N-PSDB; AA256877.
XX
PT      Recombinant DNA construct useful as vaccines for anthrax, in producing

```

host cells for analyzing the drugs and agents inhibiting anthrax -

XX PS Disclosure: Page 35; 35pp; English.

CC The invention provides a recombinant DNA construct that comprises a

CC vector and at least one nucleic acid (or its fragment) encoding a

CC combination of Bacillus anthracis proteins, selected from protective

CC antigen (PA), MAT-PA (PA with its secretory signals removed), tPA-PA

CC (PA with its secretory signals replaced with those of tissue plasminogen

CC activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine

CC for anthrax and in producing infectious alpha virus particles. These

CC particles, expressing the B. anthracis proteins are useful also as

CC vaccines for anthrax. Host cells transformed with the construct are

CC useful for analyzing the effectiveness of drugs and agents that inhibit

CC anthrax or B. anthracis proteins. The present sequence represents a

CC B. anthracis PA63 protein.

XX CC

SQ Sequence 569 AA;

Query Match 92.8%; Score 2664; DB 21; Length 569;

Best Local Similarity 96.7%; Pred. No. 2.5e-181;

Matches 521; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 20 STSAGPTVPDDNDNGIPDSLEVEGTVVYKKNRTFLSPWISIHKKKGLTKKSSPERWS 79

DB 2 stsgaptpvddndngipdslevegtyvkvknrtflspwihkkgltkykspekws 61

QY 80 TASPDPYDFEYKTVGRIDKNVSPFARHPVAAPYIVHVMENITISKNDQSTQNTDSENR 139

DB 62 tasdpdpdyfeyktvgtridknvspearhpvaapyivhvmenitliskndqstqntdsettr 121

QY 140 TISKNTSTSRHTSEVHGNAEVHANTSTSRHTSEVHGNAEVHAVAIDHSLAGERTWA 199

DB 122 tiskntstsrhtsevhgnaevhantstsrhtsevhgnaevhavaidhslagertwa 181

QY 200 ETMGINTADTARLANINRYVTGTPATRYNVLPTSTLVYLGKNOTLATIRAKENQSLIAP 259

DB 182 etmgintadtarlaninryvntgtpatrynvlpstslvylgknotlatlirakengsqiliap 241

QY 260 NNYPSKMLAPIALNAODFSSPTPTMYNOFLELEKTQRLDQYGNATATYFENG 319

DB 242 nnypskmlapialnaodfssptptmynoflelektqrlldqygnatatyfeng 301

QY 320 RVRVDTGSNMSEVLPOIQETTARILFNGKDLNVERIAAVNPSPLETPRPDMLKEAL 379

DB 302 rrvvdtgsnmsevlpoiqettarilfngkdlnveriaavnpdpлетpmdpmlkeal 361

QY 380 KIATGFNPNNGNLOYGKDTIEFPNFPQOTSQNKNOLELNATNTYTVLDTKRLNAKM 439

DB 362 kiatgfnepnngnlgygkdtiefdnfngqtsqnknglaelnaelnlytvldkrlklnakm 421

QY 440 NILRDKRFRHNRNINAVGADSVYKFAHREYINSTEGLLNTIDDKRLILSGYVEIE 499

DB 422 nilrdrkfrhndrninavgadesvykfehrevinsstegllnldkdrllksyiveie 481

QY 500 DTGEGKEVYINRYDMLNTISSLRQDGKTFIDFKKYNDKLPYLSNPKYVNVYAVTKENT 558

DB 482 dtegekevindrmylnntisslrqdgktfidfkkyndklpylsnpkyvnyvaytkent 540

RESULT 8

AAR60193

ID AAR60193 standard; Protein: 719 AA.

XX

AC AAR60193;

XX

DT 04-APR-1995 (first entry)

XX

DE Modified protective antigen of Bacillus anthracis.

XX

KM Anthrax; Bacillus anthracis; fusion protein; lethal factor;

KM protective antigen; cell killing; targeting; targeting; pathogen;

KM intracellular; HIV; human immunodeficiency virus; toxin.

XX

OS Bacillus anthracis.

XX

PN W09418332-A.

XX

PD 18-AUG-1994.

XX

PF 14-FEB-1994; 94WO-US01624.

XX

PR 12-FEB-1993; 93US-0021601.

PR 25-JUN-1993; 93US-0082849.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX

PI Arcora N, Klimpel K, Leppia SH, Nichols PJ, Singh Y;

XX

DR WPI: 1994-279753/34.

DR N-PSDB: AAQ70189.

XX

PT Nucleic acid encoding anthrax toxin fusion protein - useful for

PT targeting toxin to specific cells, eg for killing tumour cells

PT or HIV-infected cells

XX

PS Example 6; Page 114-115; 124pp; English.

XX

CC The sequence encoding the protective antigen of Bacillus anthracis

CC may be used in the construction of a nucleic acid which encodes a

CC fusion protein comprising the anthrax protective antigen binding

CC domain of the native anthrax lethal factor and a sequence encoding an

CC activity inducing domain of a second protein. Such fusion proteins

CC are useful for the specific killing of tumour cells or the killing of

CC cells infected with intracellular pathogens, especially HIV,

CC depending on the second component. The protective antigen and other

CC toxins require proteolytic cleavage to acquire activity. Since some

CC cells infected with an intracellular pathogen possess an active

CC protease with quite a narrow substrate specificity e.g. HIV, the

CC protease cleavage site found in the native toxin is replaced with an

CC intracellular pathogen specific protease site (See AAR60184-88). The

CC cleavage in cells that are infected with an intracellular pathogen

CC cleaves the modified toxin which is then rendered active and kills

CC the cell. This sequence is a modified Bacillus anthracis protective

CC antigen which has the amino acids originally at positions 162-171

CC replaced with the HIV protease cleavable sequence described in

CC AAR60186.

XX

SQ Sequence 719 AA;

Query Match 89.9%; Score 2581; DB 15; Length 719;

Best Local Similarity 90.9%; Pred. No. 2.8e-175;

Matches 511; Conservative 6; Mismatches 21; Indels 24; Gaps 2;

QY 1 DNLQLEPKOKSSNS-----RKKRSTSGAPTPVDRDNDGIPDSLEVEGTVVYKKNRTFLS 56

DB 149 dnlqlpelkqksntatlmgrnflgqtpvdrndngipdslevegtyvkvknrtfls 208

QY 57 PWISIHKKKGLTKKSSPEKMWSTADPYDFEYKTVGRIDKNVSPFARHPVAAPYIVH 116

DB 209 pwisihkkgltkykspekwmstasdpdyfeyktvgtridknvspearhpvaapyivh 268

QY 117 DMENITISKNDQSTQNTDSETRTISKNTSTSRHTSEVHGNAEVHANTSTSRHTSEVH 176

DB 269 dmenitliskndqstqntdsetrtliskntstsrhtsevhgnaevhaefldgsvsagf 328

QY 177 GNAEVHAVAIDHSLAGEKRTWAETMGINTADTARLANINRYVTGTPATRYNVLPTTS 236

DB 329 gnaevhavaidhslagertwaetmgintadtarlaninryvntgtpatrynvlpstslv 388

QY 237 LGKNNOTLATIRAKENQSLIAPNNYPSKMLAPIALNAODFSSPTPTMYNOFLELEK 296

DB 389 lgnqnotlatlirakengsqiliapnnypskmlapialnaodfssptptm----- 439

QY 297 TKOLRLDTQVYGNATYFENGVRVDTGSMNSEVLPOIETTARIIFNGKDLNLEVR 356
DB 440 -----ygnatynfengrvvdsgsmsevlpqdgetarilfngkdlnlveir 488
QY 357 IAAVNSDPLETTKPDMLKEALKAIFGNEPNGNLQYOGKIDTEPFDFNQTSQIN 416
DB 489 iaavnsdplectkpmclkaekiafgfnepgnqlgygkdlterdfndqtsqnk 548
QY 417 QLAELNATITVYLVKIKLNAMKNILIRDKRFHYDRNNIAVGADESVEKAHREVINS 476
DB 549 qlaelnatitvldkiklnakmnlirckrfhydrnlnavgadesvveahrevins 608
QY 477 EGLLINDIRKISGTYVEIEDTEGLKEVINDRYDMNISLRDQKTFIDFKKYNDK 536
DB 609 eglldirklisgtyveiedteglkevindrmdnisslrddqktdfkkyndk 668
QY 537 LPLYSNPYKVVYAVTKENT 558
DB 669 lplysnpykvnyavtkent 690
RESULT 9
AAR91239
ID AAR91239 standard; Protein; 884 AA.
AC AAR91239;
XX 14-AUG-1996 (first entry)
DT
XX B. cereus VIP1A(a) Insect-specific protein.
DE
XX Pesticide; insecticide; biological control agent; lepidoptera:
KM Coleoptera; transgenic plant; maize; insect resistance;
KW western corn rootworm; Diabrotica virgifera virgifera; VIP.
XX
OS Bacillus cereus strain AB78 (NRRL B-21058).
FH Key Location/Qualifiers
FT Peptide 1..33
FT /label= Sig_peptide
PN WO9610083-A1.
XX
PD 04-APR-1996.
XX
PF 27-SEP-1995; 95WO-EP03826.
XX
PR 05-JUN-1995; 95US-0463483.
PR 28-SEP-1994; 94US-0314594.
XX
PA (CIBA) CIBA GEIGY AG.
PI Carr B, Desai NM, Duck NB, Estruch JT, Kostichka K;
PI Koziel MG, Mullinsma, Nye GJ, Warren GW;
XX WPI: 1996-200921/20.
DR N-PSDB: AAT13940.
XX
XX Bacillus strain producing insecticidal protein during vegetative
PT growth - used in the control of lepidoptera and Coleoptera pests
XX
PS Claim 15; Page 121-124; 242pp; English.
XX
XX Insect-specific protein VIP1A(a) (AAR91239) of Bacillus cereus AB78
CC shows activity against Diabrotica spp. pests as the
CC western corn rootworm. It is encoded by the VIP1A(a) gene (AAT13940)
CC isolated from a cosmid clone of AB78. VIP1A(a) can be expressed in
CC e.g. bacterial hosts to provide biological control agents having
CC increased activity or target range, or can be expressed in transgenic
CC plants, esp. maize, to improve insect resistance. It is preferably
CC expressed as a fusion protein (see also AAR91245) with auxiliary protein
CC VIP2A(a) (AAR91238).
XX

SQ Sequence 884 AA:
Query Match 20.7%; Score 593; DB 17; Length 884;
Best Local Similarity 26.4%; Pred. No. 1e-33;
Matches 194; Conservative 95; Mismatches 213; Indels 232; Gaps 25;
QY 1 DNEQLPELKGSS-----NSRKKRSTSGAPYPPDRDNGCIPPSLEVEGTVD 47
DB 189 delrnpdfeinkesqetlakslnltfqkmkreided---tdcdgdsipdlweengytl- 244
QY 48 VKNKRTEFLSPWISNIHEKKGLTKYKSSPEKSTASPDSPDEFVGTGRIDKNAPPEARHPL 107
DB 245 ---gnrlavkwdsal-askyktkfvsnpleshvvgpdytelyekaardldsnaketnpl 300
QY 108 VAAVPIVHDMENIILSKNEDOSTQNTDSEFRTISKNTSRTHTSEVHGNAEVAHANTST 167
DB 301 vaafpsvnmekvllspnenls-----dsveshsstnw 334
QY 168 SRHTSEVHGNAEVAH-----AVAIHSLSLAGERTMAETMG---LMTADTARLNN 215
DB 335 syntegasyeaglgpkglisfgvsvnyqhselvaeq--wgtsfgntsqfntasaayltna 392
QY 216 IRYVNTGTAPIYVNLPTTSLVLGKNQTLATIKAKENQLSOLAPNNYPSKANLAPALNA 275
DB 393 vrynnvgtgaiydvkptsfvl-ndtiatitaksnstalnispgeypkkgnglaits 451
QY 276 QDDFSSTPTTMANYNQFLELEKTKQLRLDTQVYGNATYFENGVRVDTGSMNSEVLPO 335
DB 452 mdfnshpiltlnkkqvdlnlnkpmmlnetnqtdg---vykikdthgnlvggewnylq 508
QY 336 IOETTARIIFNGKDLNLEVRIRIAAVNSDPLETTKPDMLKEALKAIF--GFNEPNGNLQ 393
DB 509 ikaktlaailvdgge-rvaekrvaakdyenpedkt-psltlkdaiklsypdeikeleglly 566
QY 394 YOGKDITEFPDF--NPDQTSQINIKNOALFL-----NATNITYVLDKIKLNAMKNILIRDK 446
DB 567 yknkpilyessvmtyldentekvtqgndtqgfkdvshlydv---klpkmnvtlc-1 621
QY 447 RFHYDRNNIAVGADESVEKAHREVINSSTEG-----LLIN----- 482
DB 622 sllydn---aesndnsjgkwtntnlivesgnngkqysnnpdanltnldageklnkrd 678
QY 483 ----- 482
DB 679 yylslymkseknqtcetldgeilyptltkvvnkdknykrlidliahnksnplslhik 738
QY 483 -----IDKDIRKILSGTYVEIEDTEGL-----KEVI 508
DB 739 ndeltlfdldistdvaslpenltdeklqlysyrgikied--gllidkkgilhygefl 796
QY 509 ND-----RY-----DMNISLRDQKTFIDFKKY--NDKILPLY 540
DB 797 neasfnlepdqnytkyevysselgpnvdesklydgtkfkfdtkyskneglfy 856
QY 541 ISNPNYKVVYAVT 554
DB 857 dsqjlnwdfnnalt 870
RESULT 10
AAM60224
ID AAM60224 standard; Protein; 880 AA.
AC AAM60224;
XX
XX 28-SEP-1998 (first entry)
DT
XX Bacillus thuringiensis insecticidal toxin 177C8.
DE
XX Insecticide; pesticide; toxin; delta-endotoxin;
KW biological control; lepidopteran; coleopteran.
XX

OS	Key	Location/Qualifiers
XX	Misc-difference 253	/note= "encoded by YRA"
XX	Misc-difference 675	/note= "encoded by AC"
XX	Misc-difference 846	/note= "encoded by RAA"
XX	W09818932-A2.	
XX	07-MAR-1998.	
XX	30-OCT-1997;	97WO-US19804.
XX	30-OCT-1996;	96US-0029848.
XX	(MICO) MICOGEN CORP.	
XX	Dullum CJ, Feltelson JS, Loewer D, Muller-Cohn J,	
XX	Naraya KE, Schmelts JL, Schnepf HE, Schwab G, Stamp L;	
XX	Stoekhoff BA;	
XX	WPI: 1998-272226/24.	
XX	N-PSDB; AAV30307.	
PT	Bacillus thuringiensis isolates - used for producing pesticidal	
PT	toxins and nucleotide sequences for control of lepidopterans and	
PT	coleopterans	
PS	Claim 5; Page 81-84; 139pp; English.	
XX	This polypeptide comprises a novel soluble toxin of Bacillus	
XX	thuringiensis (B.t.) strain PS177C8 (NRRL B-21867). The toxin	
XX	belongs to a novel family of B.t. toxins that have toxicity	
XX	against non-mammalian pests. Its amino acid sequence was deduced	
XX	from a novel DNA fragment (see AAV30307) obtained by PCR from	
XX	cellular genomic DNA of PS177C8. Disclosed and claimed are novel	
XX	B.t. isolates and toxins (see AAW60218-32) that have activity against	
XX	lepidopteran and/or coleopteran pests, isolated genes, probes	
XX	and primers (see AAV30288-321 and AAV9734-87) useful for production	
XX	of the toxins and for the identification and characterisation of	
XX	these toxins, and transformed hosts, particularly plant and	
XX	bacterial hosts. The invention provides 8 entirely new families of	
XX	toxins from B.t. isolates. The toxins have the additional ability	
XX	to form pores in cell membranes, and can be used to facilitate	
XX	entry of a second agent into a target cell.	
XX	Sequence 880 AA:	
XX	Sequence 880 AA:	
XX	Query Match 20.6%; Score 591.5; DB 19; Length 880;	
XX	Best Local Similarity 29.9%; Pred. NO. 1.3e-33;	
XX	Matches 180; Conservative 96; Mismatches 219; Indels 107; Gaps 22.	
QY	1 DNLQPELKKOKSS-----NSRKKRSTSGAGPTVPDRDNGIIPDSLEVEGYTD 47	
QY	186 delinpeinfkkesgeflakpskinlftqkmkreided---tltdgdsipdlweengytl- 241	
QY	48 VKNKRTFLSPWISNIHEKKGLTKYVSSPEKWSSTASDPYSDFEKVTGRIDKNVSPARHPL 107	
QY	242 ---qnrflavkwddst-askgyklfsvnpleshtvgdpytdygaardidlnaketfnpl 297	
QY	108 VAAVYIVVDMENITLSKNEDOSTFNTOSEFTTISKNTSTSTRTHTSEVHGAHEVHANTST 167	
QY	298 vaafpsvsvmekvillspenils-----nsveshstlmw 331	
QY	168 SRTHTSEVHGAHEVH-----AAVIDHSLSLAGEPTWAEATMG---LNTQDPTARIANAN 215	
QY	332 sytlntegsgvaeiglpkpkistfsfvsvnyghsetvaeq---wgctsgntsgfntasagyllhan 389	
QY	216 IRYVNTGTAPIYVNVLPPTSLVGLKNGQTATIRAKENQOLSLAIPANNVYPSKMIATIALNA 275	

Db	390	vrynnvgcgaiydvkpticsfvl-mndiatlittaksnsalnlspgesypkkgnglats	448
Qy	276	QDDFESTPTIMYVNOFLELETKQRLDIDVOYGNIAITYNEENGRRVYDGSNMSEVLPQ	335
Db	449	moddfinsbplctlnkkqvchlnmkpmmletnqldg---vykikdchgnlvtgwgnyiqg	505
Qy	336	IQEETARIIFNGKQDNLNERRIAAVNPSDPLEFTTKPDMWTLKEALKIAF--GFNEPNCNLQ	393
Db	506	lkakastaslvdgge-rveektrvaakdyenpedkt-psltlkdaklkisyodeikeiegilly	563
Qy	334	YQGDITEFDF--NFDQOTSQINIKQIAEL----MATNITYVLDKIKLAKNMILIRDK	446
Db	564	ynknpklysswmytyidentakevtekglnndtcgkfkxvshlydv---kltpkmmvltik-l	618
Qy	447	RHHYRNMIIVAGADSSVYKEAHREVINSSTG-----LLLNID-----KQIRK	489
Db	619	sllyhn---aesndisrskwtchtnlvsgnngkkyssnmpdanlclntdaqeklnkury	675
Qy	490	ILSGYVIEDTEGLKEVYINDRYDMLNLS-SLRDQGFIFIDFKKRYNDKRLPLYSISPNYKV	548
Db	676	yalsymkseokntqceitldgeiyelptctktnvnmkndykrldiahnlk-----snplssl	730
Qy	549	NV 550	
Db	731	hl 732	
RESULT 11			
ID	AAV59277	standard; Protein: 881 AA.	
XX	AAV59277:		
XX	18-APR-2000	(first entry)	
DE	MIS	toxin from B. thuringiensis strain PS177C8.	
KW	Bacillus thuringiensis;	toxin; endotoxin; pesticidae; plant pest;	
KM	Lepidoptera;	cleopterans;	
OS	Bacillus thuringiensis.		
PH	Key	Location/Qualifiers	
FT	Misc-difference 846	/note= "unknown"	
XX	W09957282-A2.		
XX	11-NOV-1999.		
XX	06-MAY-1999;	99WO-USO0997.	
XX	06-MAY-1998;	98US-0073898.	
PA	(MYCO) MYCOGEN CORP.		
PI	Feitelson JS, Schnepf HE, Narva KE, Stockhoff BA, Schneits J;		
PI	Loewer D, Dullum CJ, Muller-Cohn J, Stamp L, Morrill G;		
PI	Finstad-Lee S;		
XX	WPI: 2000-096811/08.		
DR	N-PSDB: AAZ58786.		
PT	New polynucleotides encoding pesticidally active proteins, useful for		
PT	transforming plants for controlling pests		
XX	Disclosure: Page 68-71: 104pp: English.		
XX	The invention relates to novel B. thuringiensis isolates, and genes		
CC	encoding pesticidal toxins which are toxic to non-mammalian pests. The		
CC	genes are useful in the control of non-mammalian pests and especially		
CC	plant pests (e.g. lepidoptera and/or cleopterans). The polynucleotides		

are useful for transforming plants for controlling plant pests; for
 CC designing primers and probes useful for the identification and
 CC characterization of genes which encode pesticidal toxins. The present
 CC sequence represents a B.t. toxin.

XX Sequence 881 AA:

Query Match 20.6%; Score 591; DB 21; Length 881;
 Best Local Similarity 29.9%; Pred. No. 1.4e-33;
 Matches 180; Conservative 97; Mismatches 218; Indels 108; Gaps 22;

```

QY 1 DNQLPELKOKS-----NSRKRSTSGAPTVPRDNDGIPDSEVEGYTD 47
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 186 delrmpfinkkeqgefakpsklnlftqkmkreided---tdcdgsdpldweengytl- 241
QY 48 VKNKRFLSPWISNIEHKKGLTKYKSSPEKWSASDPYSDFEKVYGRIDKNVSPKARHPL 107
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 242 ---gnrlavkwddsl-askgytkfvsnpleshvqdydyekaaardldlsnaketfnpl 297
QY 108 VAAVPIVHVMENIILSKNEDOSTONTDSETRISKNTSRTHTSEVHGNAEVAHNTST 167
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 298 vaafpsvnmsekvllspnenls-----nveshsstlw 331
QY 168 SRTHTSEVHGNAEVAH-----AVADHSLSLAGERTMAETMG-----LNTADTARLNAN 215
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 332 syntlegasveaglgpkjglsfgvsvngqshetvage--wgtstgnstqfntasaagylman 389
QY 216 IRYVNTGTAPIYVNLPTTSLVLKKNQTLATIKAKENQSOILAPNNYPSKNLAPIALNA 275
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 390 vrynmvgtgalydvkptstfvl-nddtlatlctaksnstalnspgesyppkkgngjalts 448
QY 276 QDDFSSTPTMTMNTNOGLEKTKQLRLDQVYGNATATYFENGKRVAVDTGSWMSEVLPQ 335
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 449 mdfnshpctlnkkyvndlnlnkpmmlctnqtdg---vykikdthgnlvggengvlyqg 505
QY 336 IOETTAIRIFNGKDLNVERRIAANVPSDPLETTKPDMLTKEALKIAF--GFNEPNNQLO 393
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 506 lkaktaslivdgge-rvaekrvakdyenpedkt-psltlkdaiklsypdeikeiegllly 563
QY 394 YOGKDITEFDF--NFDQOTSQNIKNQLAEL-----NATNIYTVLDKIKLAKANNILIRDK 446
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 564 yknkplyessvmtyidentakevtkqindttgkfkdvshlydv---kltpkmvltk-1 618
QY 447 RHHYDNNINAVGADESVVKAHREVINSSTEG-----LLNLD-----KDIR 488
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 619 slllydn---aesndnsigkwtntnlvsgngngkqysnmpdanlclntdaqeklnkurd 675
QY 489 KILSGYIVEIEDTEGLKEVINDRYDMLNIS-SLRQDGKTFIDFKKYNDKPLVYSNPYK 547
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 676 yylslymksekntqceitldgeilyptlctkvnnkdknykrldliahnlk-----snplss 730
QY 548 VNV 550
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 731 lnl 733

```

RESULT 12

AAW19509 standard; Protein; 884 AA.

AAW19509;

15-APR-1998 (first entry)

B. cereus VIP1A(a) protein sequence.

Vegetative insecticidal protein: Bacillus cereus strain AB78; plant;

Insect; Sesamia nonagrioides; maize; corn borer; toxin.

Bacillus cereus.

WO9726339-A1.

```

XX 24-JUL-1997.
PD 23-DEC-1996; 96WO-EP05828.
XX 15-JAN-1996; 96GB-0000786.
XX (NOVS ) NOVARTIS AG.
XX Gay PB;
XX WPI: 1997-385342/35.
DR N-PSDB: AAT73394.
PT Protecting plants against insects of the genus Sesamia using
PT Bacillus toxic proteins - applied directly or expressed as
PT heterologous protein by the plant, also transgenic plants expressing
PT both Cry and Vip type toxins
PS Claim 8; Page 35-39; 168pp; English.
XX This is the amino acid sequence of the 100 kD vegetative insecticidal
XX protein (VIP) 1A(a) from Bacillus cereus strain AB78. The protein can
XX be used in a new method for protecting plants, and their progeny, against
XX insects of the genus Sesamia by direct or indirect application to the
XX plant (or seed or growing area). The protein is especially useful to
XX protect maize plants against the Mediterranean corn borer
XX (S. nonagrioides).
SQ Sequence 884 AA:

```

Query Match 20.5%; Score 589; DB 18; Length 884;
 Best Local Similarity 30.2%; Pred. No. 2e-33;
 Matches 180; Conservative 94; Mismatches 215; Indels 108; Gaps 22;

```

QY 1 DNQLPELKOKS-----NSRKRSTSGAPTVPRDNDGIPDSEVEGYTD 47
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 189 delrmpfinkkeqgefakpsklnlftqkmkreided---tdcdgsdpldweengytl- 244
QY 48 VKNKRFLSPWISNIEHKKGLTKYKSSPEKWSASDPYSDFEKVYGRIDKNVSPKARHPL 107
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 245 ---gnrlavkwddsl-askgytkfvsnpleshvqdydyekaaardldlsnaketfnpl 300
QY 108 VAAVPIVHVMENIILSKNEDOSTONTDSETRISKNTSRTHTSEVHGNAEVAHNTST 167
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 301 vaafpsvnmsekvllspnenls-----nveshsstlw 334
QY 168 SRTHTSEVHGNAEVAH-----AVADHSLSLAGERTMAETMG-----LNTADTARLNAN 215
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 335 syntlegasveaglgpkjglsfgvsvngqshetvage--wgtstgnstqfntasaagylman 389
QY 216 IRYVNTGTAPIYVNLPTTSLVLKKNQTLATIKAKENQSOILAPNNYPSKNLAPIALNA 275
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 393 vrynmvgtgalydvkptstfvl-nddtlatlctaksnstalnspgesyppkkgngjalts 451
QY 276 QDDFSSTPTMTMNTNOGLEKTKQLRLDQVYGNATATYFENGKRVAVDTGSWMSEVLPQ 335
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 452 mdfnshpctlnkkyvndlnlnkpmmlctnqtdg---vykikdthgnlvggengvlyqg 508
QY 336 IOETTAIRIFNGKDLNVERRIAANVPSDPLETTKPDMLTKEALKIAF--GFNEPNNQLO 393
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 509 lkaktaslivdgge-rvaekrvakdyenpedkt-psltlkdaiklsypdeikeiegllly 566
QY 394 YOGKDITEFDF--NFDQOTSQNIKNQLAEL-----NATNIYTVLDKIKLAKANNILIRDK 446
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 567 yknkplyessvmtyidentakevtkqindttgkfkdvshlydv---kltpkmvltk-1 621
QY 447 RHHYDNNINAVGADESVVKAHREVINSSTEG-----LLNLD-----KDIR 488
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 622 slllydn---aesndnsigkwtntnlvsgngngkqysnmpdanlclntdaqeklnkurd 678
QY 489 KILSGYIVEIEDTEGLKEVINDRYDMLNIS-SLRQDGKTFIDFKKYNDKPLVYSNP 544

```

Db 679 yylslymksektgcetldgeipiltktvvnkdkykrldliahnik-----snp 730

RESULT 13
AA046712
ID AA046712 standard; Protein: 884 AA.

XX
XX
AC AA046712;
XX
XX 02-JUN-1998 (first entry)
XX
DE 100 kDa VIP1A(a) toxin of *Bacillus cereus*.
XX
XX Vegetative insecticidal protein; VIP; expression; maize; protection;
KW plant; *Ostrinia furnacalis*; Asian Corn Borer; Cry toxin; VIP toxin;
KW recombinant; *Bacillus thuringiensis*; transgenic plant; resistance;
KW insect attack; *Sesamia*; maize; cereal crop.
XX
XX *Bacillus cereus*.
XX
XX WO9746105-A1.
XX
XX 11-DEC-1997.
XX
XX 27-MAY-1997; 97WO-EP02737.
XX
XX 06-JUN-1996; 96GB-0011777.
XX
XX (NOVS) NOVARTIS AG.
XX
XX Hunter B, Suwantaradon K, Utdewilligen WPM;
XX
XX WPI: 1998-041787/04.
XX
XX N-PSDB: AAV16166.
XX
XX Administration of *Bacillus* sp. toxin protein, especially Cry or
PT vegetative insecticidal protein (VIP) protein to plants - useful for
PT protection against attack by Asian Corn Borer (*Ostrinia furnacalis*)
XX
XX Claim 8: Pages 43-46; 175pp; English.

XX
XX The present sequence represents a 100 kDa vegetative insecticidal
CC protein 1A(a) (VIP1A(a)), and is derived from *Bacillus cereus* strain
CC AB78. The protein is used in a method for protecting plants and their
CC progeny against damage caused by *Ostrinia furnacalis* (Asian Corn Borer).
CC The protein is directly or indirectly applied to the plant, plant seed
CC or growing area of the plant. Cry toxins can also be used in the same
CC way, in place of VIP toxins. The Cry or VIP toxins and genes are used,
CC especially inside recombinant *B. cereus* or *B. thuringiensis* strains,
CC to produce plants protected against Asian Borer pests. Transgenic plants
CC protected against Asian Corn Borer can be used to produce seed and
CC progeny also resistant to insect attack. Plants expressing both a
CC Cry-type and a VIP toxin gene can also protect against *Sesamia* pests.
CC The method and compositions are especially used for protecting maize but
CC may also be used to protect other cereal crops against Asian Corn Borer
XX attack.
XX
XX Sequence 884 AA;
SQ

Query Match 20.5%; Score 589; DB 19; Length 884;
Best Local Similarity 30.2%; Pred. No. 2e-33;
Matches 180; Conservative 94; Mismatches 215; Indels 108; Gaps 22;

QY 1 DNLQPELKKOKS-----NSRRKRSAGPTVPDRDNDGIPDSLEVGTYVD 47
Db 189 delrnpetnkkesgeflkpsklnlftqkmkrelded---ctdgdspdlweengyl- 244
QY 48 VKNRRTFLSPWISNIHERKGLTKYKSSPEKWSSTADSPYSDPEKVTGRIDKNVSPARHPL 107
Db 245 ---qurrlavkwddsl-askgytkfsvnplshetvgdpytdyekaardldlnaketfnpl 300

QY 108 VAAPIVHVDMENTILSKNEOSTONTSETRTISKNTSTKTRHTSEVHGNAEVHANTST 167
Db 301 vaafpsvvnsmekyllspenhs-----nsveshssclnw 334

QY 168 SRTHTSEVHGNAEVH-----AAVAIDHSLAGERTAEWNG-----LNTADARLUNAN 215
Db 335 sylvtegasveaglgpkgsifgsvnyghsetvage--wgsctgntsgfnlasagylhan 392

QY 216 IRYVNTGTAPIYVNLPTTSLVYGKNOQLATIKAKENOLSQLAPNNVYPSKNLAPIALNA 275
Db 393 vrynnvgtgalydvkptcsfv1-mndlatltakenscalnspgesypkkgqgnjatts 451

QY 276 QDDFSSFTITMANYOFLLEKTKQLRLDTDOYGNIAIYNEENGVRVDTSSNMSEVLPQ 335
Db 452 mddfnshpdltnkkqvdl1nmkpmmlsetnqtdg---ykk1kdtghnltvgewngv1qg 508

QY 336 IOETARIIFNGKNDLNERRIAANVPSDPETTPKPDMLTLEALKIAF--GFNEPENGMLQ 393
Db 509 lkaktasllvdgge-rvektrvaakdyenpedkt-psltlkda1k1sydpdtkelgally 566

QY 394 YQKNDITEFDF--NEFDOOTSQNIKNQLAEL-----NATNITYVLDKIKLNKMNILARDK 446
Db 567 yknkpiyessvmtlyldentakevtqldntgtgkfkdvshlydv---kltpkmvvtlk-1 621

QY 447 RFHYDRNNIAYGADESIVYKEAHREYVINSTG-----LLNID-----KDIR 488
Db 622 sllydn---aesndns1gkctnlnvsggnmkkgysnnpdanlntldageklxnkr 678

QY 489 KILSGYIEIEDTEGLKEVINDRYDMLNIS--SLRODGKTFIDFKKYNKPLLYISNP 544
Db 679 yylslymksektgcetldgeipiltktvvnkdkykrldliahnik-----snp 730

RESULT 14
ID AAR91245
AA091245
ID AAR91245 standard; Protein: 1346 AA.

XX
XX AAR91245;
XX
XX 14-AUG-1996 (first entry)
XX
XX VIP2A(a) and VIP1A(a) fusion protein.
XX
XX Pesticide; insecticide; biological control agent; Lepidoptera;
XX Coleoptera; transgenic plant; maize; insect resistance;
XX Western corn rootworm; *Diatraea virgifera* Virgifer; VIP.
XX
XX *Bacillus cereus* strain AB89 (NRRL B-21058).
XX
XX WO9610083-A1.
XX
XX 04-APR-1996.
XX
XX 27-SEP-1995; 95WO-EP03826.
XX
XX 05-JUN-1995; 95US-0463483.
XX
XX 28-SEP-1994; 94US-0314594.
XX
XX (CIBA) CIBA GEIGY AG.
XX
XX Carr B, Desai NM, Duck NB, Estruch JF, Kostichka K;
XX Koziel WG, Mullinsma, Nye GJ, Warren GW;
XX WPI: 1996-200921/20.
XX
XX N-PSDB: AAT13944.
XX
XX *Bacillus* strain producing insecticidal protein during vegetative
XX growth - used in the control of Lepidoptera and Coleoptera pests
XX
XX Claim 30: Page 156-61.; 242pp; English.
XX
XX A fusion protein (AAR91245) is composed of the VIP2A(a) auxiliary
XX protein (AAR91238) and insect-specific VIP1A(a) protein (AAR91239) of

CC Bacillus cereus strain AB38. It is the product of a fusion gene
(AA113944) constructed from the vip coding sequences. The fusion
CC protein can be expressed e.g. in bacterial cells for use as
CC biological control agents having improved activity or target range,
CC or in transgenic plants, esp. maize, to increase insect resistance.
CC VIP1A(a) shows activity against western corn rootworm.

XX Sequence 1346 AA;

Query Match 20.5%; Score 589; DB 17; Length 1346;

Best Local Similarity 30.2%; Pred. No. 3.7e-33;
Matches 180; Conservative 94; Mismatches 215; Indels 108; Gaps 22;

```
OY 1 DNLOPELKOKS-----NSRKRSTSGPTVPDRDNGIPDSLEVEGYTD 47
DB 651 delrnpelnfkkesqeflakpsklnlftqkmkreided---tdtgdslpdlweengytl- 706
OY 48 VKNKRRTFLSPWISNIHEKKGLTKYKSSPEKSTASDPYSDFEKGVRIDKNVSPPEARHPL 107
DB 707 ---gnrlavkwddsl-askgytktfvsnplshvtdygdpytdyekaardldlsnaketfnpl 762
OY 108 VAAPIYHVDMENTILSKNEDOSTONTDSETRTISKNTSTSRHTSEVHGNAEVHANTST 167
DB 763 vaafpsvnmekvllspnenls-----nsveshstnw 796
OY 168 SRHTSEVHGNAEVH-----AVAIDHSLAGERTWAEFMG---LNTADTARLNNAN 215
DB 797 syntlegasveaglgprkqisfgvsnvygshetvage--wgtsqntsqfntasaagylan 854
OY 216 IRVYNTGTARIYVNLPTTSLVGLKNOTLATIKAKENOLSQILAPNNVYPSKNLAPILNA 275
DB 855 vrynnvgtalgaydvkptstfvl-ndtiatlaksnstalnispgeayprkqngjaits 913
OY 276 QDDFSTPTITMANTNOFLELEKTKQLRLDQVYGNATVNFENGRRVAVDTGSWSEVLPQ 335
DB 914 mddfnspitltnkqvdnllnkmmltetngtdg---vykikdthgnlvtgseengytlq 970
OY 336 IOETTAIRIIFNGKDLNVERRIAANVPSDPLETTKPDMTKEALKIAF--GFENPNGLQ 393
DB 971 ikaktaslivddge-rvaekrvaakdyenpedkt-psltlkdaiklsypdeikeleglly 1028
OY 394 YOGKDITEFDF--NFDQOTSQINIKNOIAEL-----NATNIYTVLDKIKLNKMMILIRDK 446
DB 1029 yknkprlyessvmytldentakevtkqindtqkfkdvshlydv---klprkmmvltk-1 1083
OY 447 RFHYDRNNINAVGADESIVKAEHREVINSSTEG-----LLINID-----KDIR 488
DB 1084 silydn---aesndnsigkwtntnivsgnngkqysnmpdanlntdaqeiklnkrd 1140
OY 489 KILSGYIVEIEDTEGLEKEVINDRYDMLNIS-SLRQDGKTFIDFKKYNDKLPYISNP 544
DB 1141 yylslymkseknqtcetltdgeilyptltktvnnkdnkykridliahnik-----snp 1192
```

RESULT 15
AAW19513
ID AAW19513 standard; Protein; 1346 AA.
XX
XX AAW19513;
AC
XX
XX
DT 15-APR-1998 (first entry)
DE
XX
XX B. cereus VIP1A(a)/VIP2A(a) fusion protein sequence.
XX
XX Vegetative insecticidal protein; Bacillus cereus strain AB78; plant;
KM insect; Sesamia nonagrioides; maize; corn borer; toxin; fusion protein.
XX
OS Bacillus cereus.
XX
XX W09726339-A1.
PN
XX
PD 24-JUL-1997.

```
XX  
XX 23-DEC-1996; 96WO-EP05828.  
PF  
XX  
PR 15-JAN-1996; 96GB-0000786.  
XX  
XX (NOVS ) NOVARTIS AG.  
PA  
XX  
PI Gay PB;  
XX  
XX WPI, 1997-385342/35.  
DR  
XX N-PSDB; AAT73999.  
PT  
PT Protecting plants against insects of the genus Sesamia using  
PT heterologous protein by the plant, also transgenic plants expressing  
PT both Cry and VIP type toxins  
PS  
XX Claim 8; Page 69-75; 168pp; English.  
XX  
XX This is the amino acid sequence of a fusion protein comprising the  
CC vegetative insecticidal proteins (VIP 1A(a) and VIP2A(b)) from Bacillus  
CC cereus strain AB78. The fusion protein can be used in a new method for  
CC protecting plants, and their progeny, against insects of the genus  
CC Sesamia by direct or indirect application to the plant (or seed or  
CC growing area). The protein is especially useful to protect maize plants  
CC against the Mediterranean corn borer (S. nonagrioides).
```

Query Match 20.5%; Score 589; DB 18; Length 1346;
Best Local Similarity 30.2%; Pred. No. 3.7e-33;
Matches 180; Conservative 94; Mismatches 215; Indels 108; Gaps 22;

```
OY 1 DNLOPELKOKS-----NSRKRSTSGPTVPDRDNGIPDSLEVEGYTD 47
DB 651 delrnpelnfkkesqeflakpsklnlftqkmkreided---tdtgdslpdlweengytl- 706
OY 48 VKNKRRTFLSPWISNIHEKKGLTKYKSSPEKSTASDPYSDFEKGVRIDKNVSPPEARHPL 107
DB 707 ---gnrlavkwddsl-askgytktfvsnplshvtdygdpytdyekaardldlsnaketfnpl 762
OY 108 VAAPIYHVDMENTILSKNEDOSTONTDSETRTISKNTSTSRHTSEVHGNAEVHANTST 167
DB 763 vaafpsvnmekvllspnenls-----nsveshstnw 796
OY 168 SRHTSEVHGNAEVH-----AVAIDHSLAGERTWAEFMG---LNTADTARLNNAN 215
DB 797 syntlegasveaglgprkqisfgvsnvygshetvage--wgtsqntsqfntasaagylan 854
OY 216 IRVYNTGTARIYVNLPTTSLVGLKNOTLATIKAKENOLSQILAPNNVYPSKNLAPILNA 275
DB 855 vrynnvgtalgaydvkptstfvl-ndtiatlaksnstalnispgeayprkqngjaits 913
OY 276 QDDFSTPTITMANTNOFLELEKTKQLRLDQVYGNATVNFENGRRVAVDTGSWSEVLPQ 335
DB 914 mddfnspitltnkqvdnllnkmmltetngtdg---vykikdthgnlvtgseengytlq 970
OY 336 IOETTAIRIIFNGKDLNVERRIAANVPSDPLETTKPDMTKEALKIAF--GFENPNGLQ 393
DB 971 ikaktaslivddge-rvaekrvaakdyenpedkt-psltlkdaiklsypdeikeleglly 1028
OY 394 YOGKDITEFDF--NFDQOTSQINIKNOIAEL-----NATNIYTVLDKIKLNKMMILIRDK 446
DB 1029 yknkprlyessvmytldentakevtkqindtqkfkdvshlydv---klprkmmvltk-1 1083
OY 447 RFHYDRNNINAVGADESIVKAEHREVINSSTEG-----LLINID-----KDIR 488
DB 1084 silydn---aesndnsigkwtntnivsgnngkqysnmpdanlntdaqeiklnkrd 1140
OY 489 KILSGYIVEIEDTEGLEKEVINDRYDMLNIS-SLRQDGKTFIDFKKYNDKLPYISNP 544
DB 1141 yylslymkseknqtcetltdgeilyptltktvnnkdnkykridliahnik-----snp 1192
```

Search completed: December 2, 2001, 13:49:05
Job time: 185 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:54:52 ; Search time 132.69 seconds
(without alignments)
810.236 Million cell updates/sec

Title: US-09-747-521-4_COPY_1_735
Perfect score: 3765
Sequence: 1 MKRRKVLIPMLALSTLVSS.....LYISPNKVNVAATKENT 735

Scoring table:
BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_17:*
2: sp.archaea:*
3: sp.fungi:*
4: sp.human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3649	96.9	764	2	09R0U2 bacillus an
2	3645	96.8	764	2	09R0U2 bacillus an
3	3637	96.6	764	2	09R0U2 bacillus an
4	889	23.6	876	2	032739 clostridium
5	887	23.6	876	2	09R0U2 bacillus an
6	885	23.5	879	2	09R0U2 bacillus an
7	877.5	23.3	875	2	046221 clostridium
8	798	21.2	721	2	086171 clostridium
9	211.5	5.6	204	2	09R0U2 bacillus an
10	198.5	5.3	187	2	09R0U2 bacillus an
11	194.5	5.2	3254	5	09R0U2 bacillus an
12	191.5	5.1	3130	5	09R0U2 bacillus an
13	190.5	5.1	4688	2	09R0U2 bacillus an
14	187	5.0	1072	2	09R0U2 bacillus an
15	183.5	4.9	2178	2	046149 clostridium
16	182	4.8	2647	5	09R0U2 bacillus an
17	180.5	4.8	149	2	09R0U2 bacillus an
18	179.5	4.8	962	2	049546 mycoplasma
19	179.5	4.8	1051	2	049524 mycoplasma

20	179.5	4.8	1365	2	049525 mycoplasma
21	179.5	4.8	6713	2	09R0U2 bacillus an
22	179	4.8	769	2	09R0U2 bacillus an
23	178.5	4.7	149	2	09R0U2 bacillus an
24	176.5	4.7	149	2	09R0U2 bacillus an
25	176	4.7	1193	2	045914 clostridium
26	174.5	4.6	149	2	09R0U2 bacillus an
27	174.5	4.6	4919	5	09R0U2 bacillus an
28	173.5	4.6	2269	5	026223 clostridium
29	173	4.6	1302	2	049547 mycoplasma
30	172.5	4.6	2747	5	09R0U2 bacillus an
31	172	4.6	1193	2	045914 clostridium
32	172	4.6	1939	5	025662 plasmidum
33	171.5	4.6	2529	5	025579 helicobacte
34	170.5	4.5	1837	3	074424 schizosacch
35	170	4.5	1315	2	086488 staphylococ
36	169.5	4.5	1621	5	024984 giardia lam
37	168.5	4.5	2771	5	026216 plasmidum
38	168	4.5	2522	5	077365 plasmidum
39	168	4.5	3724	5	077320 plasmidum
40	167.5	4.4	1176	2	09R231 helicobacte
41	166.5	4.4	1115	3	094488 schizosacch
42	166.5	4.4	1230	2	025772 helicobacte
43	166.5	4.4	1344	2	049545 mycoplasma
44	166	4.4	604	5	026021 plasmidum
45	166	4.4	821	3	008581 saccharomyc

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	764 AA.
09R0U2	09R0U2			
AC	01-MAY-2000 (TREMURel. 13, Created)			
DT	01-MAY-2000 (TREMURel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMURel. 17, Last annotation update)			
DE	px01-110 (PROTECTIVE ANTIGEN).			
GN	PAG.			
OS	Bacillus anthracis.			
OG	Plasmid virulence plasmid px01, and plasmid px01.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1392;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=STERNE; PLASMID=VIRULENCE PLASMID PX01;			
RX	MEDLINE=9445483; PubMed=10515943;			
RA	Okinaka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hill K.K.,			
RA	Keim P., Koehler T.M., Lanke G., Kumano S., Mahillon J., Manter D.,			
RA	Marlnez Y., Rieke D., Svensson R., Jackson P.J.,			
RT	"Sequence and organization of px01, the large Bacillus anthracis			
RT	plasmid harboring the Anthrax toxin genes."			
RL	J. Bacteriol. 181:6509-6515(1999).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=33, 28, AND BA1035; PLASMID=PX01;			
RC	Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;			
RA	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.			
RL	EMBL: AF065404; AAD32414.1; -			
DR	EMBL: AF06781; AAG24449.1; -			
DR	EMBL: AF306778; AAG24448.1; -			
DR	EMBL: AF306779; AAG24447.1; -			
DR	EMBL: AF306780; AAG24448.1; -			

DR HSP: P13423; IACC.
DR InterPro: IPR003896; Binary_toxB.
DR PRINTS: PR01391; BINARYTOXB.
KW Plasmid.
SQ SEQUENCE 764 AA; 85810 MW; 3AE1EFBA48FA03F CRC64;

Query Match 96.9%; Score 3649; DB 2; Length 764;
Best Local Similarity 97.4%; Pred. No. 4.7e-204;
Matches 716; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

OY 1 MKRRVLPPLMALSTIVSTGNLEVIAEQVKEOENRLNESSSSGGLGYYFSDLNFOA 60
DB 1 MKRRVLPPLMALSTIVSTGNLEVIAEQVKEOENRLNESSSSGGLGYYFSDLNFOA 60
OY 61 PMVVTSTTGDISIPSSSELENIPSENOYFQSAIWSGFIKKKSDTEYFATSADNHTMYV 120
DB 61 PMVVTSTTGDISIPSSSELENIPSENOYFQSAIWSGFIKKKSDTEYFATSADNHTMYV 120
OY 121 DDOEVINKASNSNKIRLEKGRLYQIKYQRENPTREKGLDFKLYWTDSONKKEVISSDNL 180
DB 121 DDOEVINKASNSNKIRLEKGRLYQIKYQRENPTREKGLDFKLYWTDSONKKEVISSDNL 180
OY 121 DDOEVINKASNSNKIRLEKGRLYQIKYQRENPTREKGLDFKLYWTDSONKKEVISSDNL 180
DB 121 DDOEVINKASNSNKIRLEKGRLYQIKYQRENPTREKGLDFKLYWTDSONKKEVISSDNL 180
OY 181 QLPBCLKOKSSNRKRRSTASGPTVPDRNDGIPDSLEVEGYVDVKNKRTFLSPWISNTH 240
DB 181 QLPBCLKOKSSNRKRRSTASGPTVPDRNDGIPDSLEVEGYVDVKNKRTFLSPWISNTH 240
OY 241 EKKGLTKYKSSPEKWSASDPYSDEKVTGRIDKNVSPRARHPLVAAPYIVHVDMENIL 300
DB 241 EKKGLTKYKSSPEKWSASDPYSDEKVTGRIDKNVSPRARHPLVAAPYIVHVDMENIL 300
OY 301 SKNEOSTONTSETRTISKNTSTSRTHSEVHGNAEVAHANTSTSRTHSEVHGNAEVAH 360
DB 301 SKNEOSTONTSETRTISKNTSTSRTHSEVHGNAEVAHANTSTSRTHSEVHGNAEVAH 360
OY 361 VAIDHSLSLAGRTAETGTLTADTARLANIRVYNTGTAPIYVLPPTSVLGKNQTL 420
DB 361 VAIDHSLSLAGRTAETGTLTADTARLANIRVYNTGTAPIYVLPPTSVLGKNQTL 420
OY 421 ATIKAKENQLSQILAPNNYPSKNLAPIALNADDFSSPTITMANYNOFLEKTKQLRLD 480
DB 421 ATIKAKENQLSQILAPNNYPSKNLAPIALNADDFSSPTITMANYNOFLEKTKQLRLD 480
OY 481 TPOVGNATYTFENGRAVVDGSMSEVLPQIOETTAIRIENGKDLNVERRIAIVNPS 540
DB 481 TPOVGNATYTFENGRAVVDGSMSEVLPQIOETTAIRIENGKDLNVERRIAIVNPS 540
OY 541 DPLETTKPDMTLEALKIAGFENEPNGNLQYOGKDIPEDFNPDOTSONIKNOIAELNA 600
DB 541 DPLETTKPDMTLEALKIAGFENEPNGNLQYOGKDIPEDFNPDOTSONIKNOIAELNA 600
OY 601 TNIYTVLDKIKLNAKMNLIRDKRPHYDRNNIYAVGADSVKAEHREVINSSTEGILLNI 660
DB 601 TNIYTVLDKIKLNAKMNLIRDKRPHYDRNNIYAVGADSVKAEHREVINSSTEGILLNI 660
OY 661 DKDIIKILSGYIVETEDTGLKEVINDRDMLNISLRDQGTFTDFKKNKPLIYISN 720
DB 661 DKDIIKILSGYIVETEDTGLKEVINDRDMLNISLRDQGTFTDFKKNKPLIYISN 720
OY 721 PMYKVNVAVTKEKT 735
DB 721 PMYKVNVAVTKEKT 735

RESULT 2

O9KH69 PRELIMINARY; PRT; 764 AA.

AC O9KH69;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PROTECTIVE ANTIGEN.
GN PACA OR PAG.

OS Bacillus anthracis.
OG Plasmid pX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V770-NP1-R, ATCC14185, PLASMID-PX01;
RA MEDLINE=20359347; PubMed=10899854;
RA Cohen S., Mendelson I., Alboum Z., Kohler D., Elhanany E., Bino T.,
RA Leitner M., Inbar I., Rosenberg H., Gozes Y., Barak R., Fisher M.,
RA Kromman C., Velan B., Shaffer A.,
RT "Attenuated nontoxicogenic and nonencapsulated recombinant Bacillus
RT anthracis spore vaccines protect against anthrax".
RL Infect. Immun. 68:4549-4558(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC PLASMID-PX01;
RA MEDLINE=99214082; PubMed=10197996;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RT "Genetic diversity in the protective antigen gene of Bacillus
RT anthracis".
RL J. Bacteriol. 181:2358-2362(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC PLASMID-PX01;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF268967; AAF6457.1; -
DR EMBL: AF306782; AAG24450.1; -
DR InterPro: IPR003896; Binary_toxB.
DR PRINTS: PR01391; BINARYTOXB.
KW Plasmid.
SQ SEQUENCE 764 AA; 85838 MW; B0DAFC1DC83DAF4 CRC64;

Query Match 96.8%; Score 3645; DB 2; Length 764;
Best Local Similarity 97.3%; Pred. No. 8.1e-204;
Matches 715; Conservative 3; Mismatches 17; Indels 0; Gaps 0;

OY 1 MKRRVLPPLMALSTIVSTGNLEVIAEQVKEOENRLNESSSSGGLGYYFSDLNFOA 60
DB 1 MKRRVLPPLMALSTIVSTGNLEVIAEQVKEOENRLNESSSSGGLGYYFSDLNFOA 60
OY 61 PMVVTSTTGDISIPSSSELENIPSENOYFQSAIWSGFIKKKSDTEYFATSADNHTMYV 120
DB 61 PMVVTSTTGDISIPSSSELENIPSENOYFQSAIWSGFIKKKSDTEYFATSADNHTMYV 120
OY 121 DDOEVINKASNSNKIRLEKGRLYQIKYQRENPTREKGLDFKLYWTDSONKKEVISSDNL 180
DB 121 DDOEVINKASNSNKIRLEKGRLYQIKYQRENPTREKGLDFKLYWTDSONKKEVISSDNL 180
OY 121 DDOEVINKASNSNKIRLEKGRLYQIKYQRENPTREKGLDFKLYWTDSONKKEVISSDNL 180
DB 121 DDOEVINKASNSNKIRLEKGRLYQIKYQRENPTREKGLDFKLYWTDSONKKEVISSDNL 180
OY 181 QLPBCLKOKSSNRKRRSTASGPTVPDRNDGIPDSLEVEGYVDVKNKRTFLSPWISNTH 240
DB 181 QLPBCLKOKSSNRKRRSTASGPTVPDRNDGIPDSLEVEGYVDVKNKRTFLSPWISNTH 240
OY 241 EKKGLTKYKSSPEKWSASDPYSDEKVTGRIDKNVSPRARHPLVAAPYIVHVDMENIL 300
DB 241 EKKGLTKYKSSPEKWSASDPYSDEKVTGRIDKNVSPRARHPLVAAPYIVHVDMENIL 300
OY 301 SKNEOSTONTSETRTISKNTSTSRTHSEVHGNAEVAHANTSTSRTHSEVHGNAEVAH 360
DB 301 SKNEOSTONTSETRTISKNTSTSRTHSEVHGNAEVAHANTSTSRTHSEVHGNAEVAH 360
OY 301 SKNEOSTONTSETRTISKNTSTSRTHSEVHGNAEVAHANTSTSRTHSEVHGNAEVAH 360
DB 301 SKNEOSTONTSETRTISKNTSTSRTHSEVHGNAEVAHANTSTSRTHSEVHGNAEVAH 360
OY 361 VAIDHSLSLAGRTAETGTLTADTARLANIRVYNTGTAPIYVLPPTSVLGKNQTL 420
DB 361 VAIDHSLSLAGRTAETGTLTADTARLANIRVYNTGTAPIYVLPPTSVLGKNQTL 420
OY 421 ATIKAKENQLSQILAPNNYPSKNLAPIALNADDFSSPTITMANYNOFLEKTKQLRLD 480
DB 421 ATIKAKENQLSQILAPNNYPSKNLAPIALNADDFSSPTITMANYNOFLEKTKQLRLD 480
OY 481 TPOVGNATYTFENGRAVVDGSMSEVLPQIOETTAIRIENGKDLNVERRIAIVNPS 540


```
Db 481 TDQVYGNIAATYFNENGVRVDTGSMNSEVLPQIOETARIIFNGKDLNVERRIAANPS 540
Qy 541 DPLETTKPDMTLKEALKIAGFNEPNCNLOYOGKDIETEPFNFDOQTSONIKNOELAE 600
Db 541 DPLETTKPDMTLKEALKIAGFNEPNCNLOYOGKDIETEPFNFDOQTSONIKNOELAE 600
Qy 601 TNYTYVLDKIKLNKMMILIRDKRFHYDRNNIAVGADESIVKEAHEVINSSTEGILLNI 660
Db 601 TNYTYVLDKIKLNKMMILIRDKRFHYDRNNIAVGADESIVKEAHEVINSSTEGILLNI 660
Qy 661 DKDIRKILSGYIEIEDTEGLKEVINDRYDMLNISSLRQDKTFIDFKKYNDKLPYISN 720
Db 661 DKDIRKILSGYIEIEDTEGLKEVINDRYDMLNISSLRQDKTFIDFKKYNDKLPYISN 720
Qy 721 PNKYVNVYAVTKENT 735
Db 721 PNKYVNVYAVTKENT 735

RESULT 3
Q9FSR7 PRELIMINARY; PRT; 764 AA.
AC Q9FSR7;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE PROTECTIVE ANTIGEN.
GN PAG.
OS Bacillus anthracis.
OC Plasmid pX01.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Bacillus.
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BA1024;
RX MEDLINE-99214082; PubMed-10197996;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RT "Genetic diversity in the protective antigen gene of Bacillus anthracis."
RT J. Bacteriol. 181:2358-2362(1999).
[2]
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN-BA1024;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF306783; AAG24451.1; -.
DR InterPro; IPR003896; Binary_toxb.
DR PRINTS; PR01391; BINARVTOXINB.
KW plasmid.
SQ SEQUENCE 764 AA; 85828 MW; A1845CE1FEDCD93A CRC64;
```

Query Match 96.6%; Score 3637; DB 2; Length 764;
Best Local Similarity 97.1%; Pred. No. 2, 4e-203;
Matches 714; Conservative 3; Mismatches 18; Indels 0; Gaps 0;

```
Qy 241 EKKGLTKYKSSPEKWSASDPYSDFEKTGTGRIDKNSPEARHPLVAAVPIVHVMENITL 300
Db 241 EKKGLTKYKSSPEKWSASDPYSDFEKTGTGRIDKNSPEARHPLVAAVPIVHVMENITL 300
Qy 301 SKNEDQSTQMTDSEPTISKNTSFRTHTESEVHNAEVAHNTSTSRTHTESEVHNAEVAH 360
Db 301 SKNEDQSTQMTDSEPTISKNTSFRTHTESEVHNAEVAHNTSTSRTHTESEVHNAEVAH 360
Qy 361 VAIDHSLSLAGEKRTWAETMGLNTADTARLNANIRYVNTGTAPIYVNLPTTSLVIGKNOTL 420
Db 361 VAIDHSLSLAGEKRTWAETMGLNTADTARLNANIRYVNTGTAPIYVNLPTTSLVIGKNOTL 420
Qy 421 ATTAKKENQSLIAPANNYPSPKPLAPALNAODFSSPTPTMNVNPFLEKTKOIRLD 480
Db 421 ATTAKKENQSLIAPANNYPSPKPLAPALNAODFSSPTPTMNVNPFLEKTKOIRLD 480
Qy 481 TDQVYGNIAATYFNENGVRVDTGSMNSEVLPQIOETARIIFNGKDLNVERRIAANPS 540
Db 481 TDQVYGNIAATYFNENGVRVDTGSMNSEVLPQIOETARIIFNGKDLNVERRIAANPS 540
Qy 541 DPLETTKPDMTLKEALKIAGFNEPNCNLOYOGKDIETEPFNFDOQTSONIKNOELAE 600
Db 541 DPLETTKPDMTLKEALKIAGFNEPNCNLOYOGKDIETEPFNFDOQTSONIKNOELAE 600
Qy 601 TNYTYVLDKIKLNKMMILIRDKRFHYDRNNIAVGADESIVKEAHEVINSSTEGILLNI 660
Db 601 TNYTYVLDKIKLNKMMILIRDKRFHYDRNNIAVGADESIVKEAHEVINSSTEGILLNI 660
Qy 661 DKDIRKILSGYIEIEDTEGLKEVINDRYDMLNISSLRQDKTFIDFKKYNDKLPYISN 720
Db 661 DKDIRKILSGYIEIEDTEGLKEVINDRYDMLNISSLRQDKTFIDFKKYNDKLPYISN 720
Qy 721 PNKYVNVYAVTKENT 735
Db 721 PNKYVNVYAVTKENT 735

RESULT 4
ID 032739 PRELIMINARY; PRT; 876 AA.
AC 032739;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE ADP-RIBOSYLTRANSFERASE.
GN CDTB.
OS Clostridium difficile.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1496;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CD196;
RX MEDLINE=97230316; PubMed=9119480;
RA Perelle S., Gilbert M., Bourlioux P., Corthier G., Popoff M.R.;
RT "Production of a complete binary toxin (actin-specific ADP-ribosyltransferase) by Clostridium difficile CD196."
RT Infect. Immun. 65:1402-1407(1997).
DR EMBL; U76081; AAB67305.1; -.
DR HSSP; P13423; IACC.
DR InterPro; IPR003896; Binary_toxb.
DR PRINTS; PR01391; BINARVTOXINB.
KW transferase.
SQ SEQUENCE 876 AA; 98796 MW; 25E06E2D45CE2B3B CRC64;
```

Query Match 23.6%; Score 889; DB 2; Length 876;
Best Local Similarity 32.5%; Pred. No. 1, 3e-43;
Matches 261; Conservative 139; Mismatches 274; Indels 128; Gaps 34;

DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JAN-1998 (Tremblrel. 05, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE SB COMPONENT.
GN SBL.
OS Clostridium spiroforme.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes.
OX NCBI_Taxid=29348;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CS246;
RA Gilbert M., Perelle S., Daube G., Popoff M.R.;
RL Syst. Appl. Microbiol. 20:337-347(1997).
DR EMBL; X97699; CAA66612.1; -.
DR HSSP; P13423; IACC.
DR Interpro: IPR003896; Binary_toxB.
DR PRINTS; PRO1391; BINARYTOXINB.
SQ SEQUENCE 879 AA; 98738 MW; 40685ACB8E05BA01 CRC64;

Query Match 23.3%; Score 885; DB 2; Length 879;
Best Local Similarity 29.0%; Pred. No. 2, 2e-43;
Matches 267; Conservative 148; Mismatches 269; Indels 236; Gaps 36;

QY 1 MKRKVLPDMLSTVLSTGNTLEVIQAEV-----KQENRLNSESSESSOGLLGYFSD 55
DB 1 MNKRLTGLTLC--TVVGOMMTYPIYAKTIIONNOGEVETINERTKVSNGMLGYFEAD 58
QY 56 LNFQAPMVYTSSTGDLSPSSLENIPEBNQ-YFQSAIWSGFYKYSDEYFATSADN 114
DB 59 EHFKEDELKAPYKNGELKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEK 118
QY 115 HTVMWDDDEVINKANSNKKIRLEKRLQIKYQORENTEKGLDF-----KLYMTD 167
DB 119 -VLMQINAE--GEIANTPLKVNKKGQEVIRLEIQ---DKDIGVDDLSPKLYW-E 168
QY 168 SONKEVSSDNLQELPELKQSSN-----SRKRSTSA-GPYVPRDNDG 211
DB 169 LMGDKTLPIEKNFLRDYKIDENDFIRKDNFEDLKLSRSARLASGWCDELDLDNDN 228
QY 212 IPDSLEVEGYVDVKNKRKTFPLSPWISNIHEKKGLTYKSSPEKWSGASPSDFEYVGR 271
DB 229 IPDAYEKNQYTI---KDSIAVYK-W-EDSFAQGGYKKRYLSSYSLESTNAGDPYDYQKASG 283
QY 272 IOKNVSPEARHPLAAYPIYHVDMENILISKNDOSTQNTDSTRTISKTSRTP--HT 329
DB 284 FKAIAEAARDPLAAYPVAVGVGMKLIISTNEHAST---DQKTYVSRTNTSKTDANT 339
QY 330 SEVHGNAEVH---ANTSTSRTHSEVHGNAEVHAVAIDHSLSLAGEPRTAETMGILTAD 385
DB 340 AGVAINIAVONGFTGSIITNYSHTTE-----NSTAVONS---NGE-SMNTSLSTIKGE 388
QY 386 TARLANINRYVTGTAPRIYVNLPTSTVLGKNOTLATIRAKENQLSQIILANNRYESK 445
DB 389 SAYINANRYVTGTAPRIYVNLPTSTVLGKNOTLATIRAKENQLSQIILANNRYESK 447
QY 446 APLALNAODESSPTITMANYNGFLELEKTKQLDLDYVYGNATATNFENGRRVDT-GS 504
DB 448 SPLALNTMOQFSSRLPIPNYDOLKIDAGKQIKLETTVOVSGN---YGIKNSQOQIITEGN 504
QY 505 MNSEVLPOIOWETARIIFN-GKDLNLVERIAVNPSPDEPPTTKPMTLKEAKIKFGEN 563
DB 505 SMSDYISQDLSASITLDTGSD--VFERRVYAKDSSNPEDKT-PVLTIGEALEKKAFGAT 561
QY 564 EFGNGLQYQKQDITE--FDENFDQOSTONIKNOJLAELNATNITYVLDKIKLANMILIR 621
DB 562 KNGEILYFGMPIDESCVELIFDGNANTLIKERLNLNDKIKYIV---QLEGMKILIK 617
QY 622 D----- 622
DB 618 TSTYFNNDGYNFNPSSWSNVDSONODGLONAMANKLGETKIVIPSKLNPYKRYVFSGY 677
QY 623 -----KRPHY-----DRNN----- 631

DB 678 LKNSSTNPITVNIKAKEQTYNLVSENDYKKKSEYEFETIGRNASNIETILTSSGITFLD 737
QY 632 -----IANGADESVYKFAHREV-----INSTEBLLN-----I 660
DB 738 NLSITELNSTPELKEPDIKVPSPDQELI-DAHKRYVADLSFNOSTANVYLDGLYFEPQOT 796
QY 661 DKDIRKLSGYIYEIE-DTEGLKEVINDRYDMNISSLRDQGT-FIDFKYNDKLPYLI 718
DB 797 NKEVLDYQKRYKVEATILEYSGFKDIGTKKELRNYTDSNQPKTINYNFYSY-----FT 850
QY 719 SNPNY-----KVNYAVATKEN 734
DB 851 SGENVMPYKKRLRIYATIPEN 870.

RESULT 7
ID 046221
AC 046221; PRELIMINARY; PRT; 875 AA.
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE IOTA TOXIN COMPONENT IB PRECURSOR.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OX NCBI_Taxid=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCIB 10748;
RX MEDLINE=94041637; PubMed=8225592;
RA Perelle S., Gilbert M., Bogue P., Popoff M.R.;
RT "Characterization of Clostridium perfringens Iota-toxin genes and
expression in Escherichia coli.";
RL Infect. Immun. 61:5147-5156(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-NCIB 10748;
RA Popoff M.R.;
RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-NCIB 10748;
RA Popoff M.R.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X73562; CAA51960.1; -.
DR HSSP; P13423; IACC.
DR Interpro: IPR003896; Binary_toxB.
DR PRINTS; PRO1391; BINARYTOXINB.
KW SIGNAL.
FT SIGNAL. 34 38
FT CHAIN 212 875 IOTA TOXIN COMPONENT IB.
SQ SEQUENCE 875 AA; 98468 MW; C9AE092CD3818921 CRC64;

Query Match 23.3%; Score 877.5; DB 2; Length 875;
Best Local Similarity 31.9%; Pred. No. 5, 8e-43;
Matches 254; Conservative 132; Mismatches 270; Indels 139; Gaps 33;

QY 15 TLIVSSGNTLEV-----IOAEVQENRLNSESSESSOGLLGYFSDNFPQAPMVYTSST 69
DB 15 TAMIQSLSTYNVAQTTQNDTNOKEETINENTLSSGMLGYFADHEFDLEIMADIKN 74
QY 70 GDLSPSSSELENIPESE-NOYFQSAIWSGFYKYSDEYFATSADNHYVTMWVDDQEVYINK 128
DB 75 GDLKPEEKYVDKLTLEDNNSIKSIKRWGTRIIPSDGYILSTDR-NOYVLMQINAKGDIK 133
QY 129 ASNSNKRLEKGRLYQIKIYOREN-----PTKGLDFKLYWTDSONKKEVSSDN 179
DB 134 ---TLKVNMRKGQAYNRIEIODKNLGSDIDLNSVP-----KLYW-ELMGNKVYIPEEN 182
QY 180 LQELPELKQSSNSKRRKSTSAQTPV-----DNDNGIDPSLEVEG 220

```
Db 183 LPRFRYSKIDEDD-----PPIPNNNPFQVRFSSAMEDEDDIDTDNDNIPDAVERNG 233
Qy 221 YTVVKKNRKRLSPWISNIHEKKGLTKYKSPKSTASDPYDEKATGRIDKVVSPA 280
224 YTI-----KDSIAVKMNDSEFAE-QGYKKVYSSLESNTAGDPYTDVQKSGSIDKAIKLEA 288
Qy 281 RHPLYAAPIYHVDMENTILSKNEDOSTQNTDSETRTISKNTSRTSRTSEVHGNAEYHA 340
289 RDLPLAAAPVYGVMENILISTNEHASS-----DQKTVSRATNTSKNTANTV--GVSISA 342
Qy 341 -----NTGSRTRTSEVHGNAEYHVAIDHSLSLAGEPTMATMGLNADTARLAN 392
343 GYQNGFTGNITTSYSHTD-----NSTAVODS---NGE-SMWTGLSINGESAYIVAN 391
Qy 393 IRYVNTGTAPIYVNLPTTSYLVKKNQTLATIKAKENQLSQILAPNNYPSKLIAPIALNA 452
392 VRYNTGTAPYKATPTNTLV-DEGTIATIKADONQIGNNISPREYTPKKGSLPLAINT 450
Qy 453 ODDFSSPTTNNYNOFLEKTKOLRLDTPDYGNIAATYFENGVRVDTGSSNMSEVLPO 512
451 MDOFNARLIPINYDOLKLDGSKQIKLETTQVSGNYGTKN--SQOI-IIEGNSMGNYISQ 508
Qy 513 IQETARIRFNGKOLNVERIAANPSDPLETTPKPMTLKALKIAGFNEPNKLOY- 571
509 IDSASASILLD-TGSQTFERRVAAKEQGNPEDKT-PEITIGELAKKASATK-NEGLLYF 565
Qy 572 QGKDITE--PDFNEQDOTSONIKNOELANATNTYTVLDKIKLNKMLIRDRKF--HY 627
566 NGIPIDECVELIFDDNSELKEIKELKYLDDKAIYV----KLENGMILLIKVPSPYTFNF 621
Qy 628 DR-NNIAVGADESVYKKAHREYINSTEGL-----LNIIDKIRKILSGY----- 671
622 DEYNNF--ASMSNIDTQNDQLOSVANKRLSGEFKIIIPMSKIKPKRYKRVFSGSKDPST 679
Qy 672 -----IVEIEDREGLEKVENDRYDMLNIS-----SLROGKTFIDPKKIND 712
680 SNSTVNIKSQKTDVYVPEKDYTKFSYEFETTGKSSDIETTLTSSGVIFLDMLSITE 739
Db 713 --KLPLYSINPNKY 725
740 LNSTPELKEPEIKV 754
RESULT 8
086171 PRELIMINARY; PRT; 721 AA.
AC 086171
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE C2 TOXIN (COMPONENT-11).
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TYPE C (C)-203028;
RX MEDLINE=98323874; PubMed=9659689;
RA Kimura K., Kubota T., Ohishi I., Isogai H., Isogai E., Fujii N.;
RT "The gene for component-II of botulinum C2 toxin.";
RL Vet. Microbiol. 62:27-34(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-TYPE C (C)-203028;
RX MEDLINE=96184657; PubMed=8645309;
RA Fujii N., Kubota T., Shiraoka S., Kimura K., Ohishi I., Morishiki K.,
RA Isogai E., Isogai H.;
RT "Characterization of component-I gene of botulinum C2 toxin and PCR
RT detection of its gene in clostridial species.";
RL Biochem. Biophys. Res. Commun. 220:353-359(1996).
DR EMBL; D88982; BAA32537.1; -.
```

```
DR HSSP; P13423; IACC.
DR InterPro; IPR003896; Binary_toxb.
DR PRINTS; PRO1391; BINARYTOXINB.
SQ SEQUENCE 721 AA; 80515 MW; 44C8153AC749D5F2 CRC64;

Query Match 21.2%; Score 798; DB 2; Length 721;
Best Local Similarity 34.5%; Pred. No. 1,86-38;
Matches 221; Conservative 117; Mismatches 221; Indels 82; Gaps 25;

Qy 32 KOENLNESES--SSOGLIGYFSDLFQOAPMVYTSSTDDLSIPSELEINIPSENYF 89
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 5 KPENSVKNNKNYFTNLGMLGYFEN-DFENLNTISPLDGLNLFKEDINSTIG-NRIT 62
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 90 QSAISGFIKVKKSDGYFFATASDN-HVTMWVDDQEVINKASN--SNKIRLEKGRLOYIKI 147
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 63 KSARIGILKRSIGCEIYLLTNSPCRYEL---NEIRNLSTNSNTYVNLQGNAYDIRI 119
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 148 -QYORENPTEKGLD-FKLYWTDSONKKEVISSDNLQLELAKOKSSNRKSTSGPTV- 204
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 120 EQLMSENOLLKNYEGIKLYWETSDIIRKELIPSEVL---LKPNSYNTNEKSKFIPNNTLF 175
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 205 -----PDRNDGIPDLSIEVGYVDYKKNRTFELSPMISNIHEKKGLTKYKSSPEK 254
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 176 SNAKIKANANDTDGDIPEDEMEINGTYVMNQAVAMDCKEAA--GKKKYVSNPK 230
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 255 WSTASDPYSDFEKYGRIDKNVSPARHPLVAAPIYHVDMENTILSKNEDOSTQNTDSE 314
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 231 PCTANDPYTDEKYSQGLDPSVSMVARDPMISATYIVGVOMERLVKSE---YITDSD 286
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 315 TRTISKNTSRTHT---SEVHGNAEV-----HANTSTSKTHTSEVHGNAEV 358
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 287 TKSMKSTSHSTNTNITGAEVSGSLQAGIFPVFSMASANSYHTQONTSTVD----- 341
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 359 HAVAIHSLSLAGEPTMATMGLNADTARLANRNYNTGTAPIYVNLPTTSYLVKNO 418
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 342 -----DTTGE-SFSQGSJINGESAYINPNIRYNTGTAPIYVNLPTTYVIDK-Q 390
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 419 TLATIKAKENQLSQILAPNNYPSKNLAPIALNADDDSPITMNNYNOFLEKTKOLR 478
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 391 SVATIKGOESLIGDYLNNGGYTPIIGEPFMALNTMDQSSRLPIPNYNOGLKSIDGCTVM 450
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 479 LPTDOYVGNIAATYFENGVRVDTGSSNMSEVLPOIETARTI--FNGKDLNVERRIA 536
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 451 LSTSQFTGNFAKYN--SNGLNLYTD--GNMNGPYLGITIKSTYASLTSFSQOTQVA---VVA 505
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 537 VNPSDPLETTPKPMTLKALKIAGFNEPNKNGNLQOGQDIT---EFDNPFQDOTSONIKN 593
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 506 PNFSDPEDKT-PKLTLEQALYKAFALKKNGKFFHGLEISKNEKIOVFLDSNTNDFEN 564
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Qy 594 QLAELNATNTYTVLDKIKLNKMLNIRDRFHYDRNIAV 634
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |
Db 565 QLKNTADKDINHCI--IKRN--MNLIVAVITFKENISSINI 601
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : |

RESULT 9
09X377 PRELIMINARY; PRT; 204 AA.
AC 09X377
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, last annotation update)
DE PX01-111.
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-STERNE;
RA Ohtnaka R.T., Cloud G., Hamton O., Hoffmaster A., Hall K.K., Keim P.,
RA Koehler T., Lamke G., Kumano S., Mahillion J., Mantler D., Martinez Y.,
```

RA Ricke D.O., Svensson R., Jackson P.J.;
RT "Type sequence and organization of pXO1, the large *Bacillus anthracis*
RT plasmid harboring the anthrax toxin genes."; [J. Bacteriol.](#) 0:0-0(1999).
RL EMBL: AF065404; AAD32415.1; -.
DR HSSP; P13423; IACC.
KW Plasmid.
SQ SEQUENCE 204 AA; 23029 MW; E1657B23AE4273FD CRC64;

Query Match	5.6%;	Score 211.5;	DB 2;	Length 204;
Best Local Similarity	34.8%;	Pred. NO. 3.7e-05;		
Matches 46;	Conservative 31;	Mismatches 34;	Indels 21;	Gaps 4

```

OY 616 MNILIRDRFRFEDNNINIAVDESVEAEHAEVINSSTEGGLINDIDOKRISSTGYVEI 675
Db 1 MNILVRDP-YHYDDNNGNVGVDSDYLKNAKQJLNMWSSDDGVSLSNLTDEDAVQALSGYMLQI 59
OY 676 EDTE-----GLKEVINDRYDMINISLRDGGKTFIDEKKYVINKPLXYINPN 722
Db 60 KKPSSHLITNSPVTTITLACKDQSGVELRYVLS-----DGTGFLDENKFKDENWMSLV-DPG 112
OY 723 YKANVYATVKEN 734
Db 113 DDVYVYATVKED 124

```

SEQUENCE	1387 AA	164871 MW	5660154744FC0285 CRC64
SO	NON_TER	1	1
FT	1387	1387	
DR	EMBL: AF196347; AAF98066.1; -		
RL	Proc. Natl. Acad. Sci. U.S.A. 97:9648-9653(2000).		
RT	Involved in host cell selection and invasion."		
RT	related to Plasmodium vivax and Plasmodium yoelii adhesive proteins		
RA	Ravner J.C., Galinski M.R., Ingravallo P., Barnwell J.W.		
RX	MEDLINE=20402589; PubMed=10920203;		
RP	SEQUENCE FROM N.A.		
OC	NCB1_TaxID=5833;		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
DE	RETICULOCYTE-BINDING PROTEIN 2 HOMOLOG A (FRAGMENT).		
OS	Plasmodium falciparum.		
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)		
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)		
AC	096276;		
ID	096276	PRELIMINARY;	PRT; 1387 AA.
RESULT	10		

Query Match	5.38;	Score 198.5;	DB 5;	Length 1387;
Best Local Similarity	19.68;	Pred. No. 0.0031;		
Matches 169;	Conservative 136;	Mismatches 289;	Indels 269;	Gaps 41

OY	5	KVPELMASTIIIVASNGNIEVIOAQEAKOEORRLINBESSQGLGYFDLNGQANV	64
Db	284	KHVIKLANFSIIIVSDTNETITPENPLEDDLN	329
OY	65	TSSTTGDISSPSSLENI--PSENOYFOSAIWGSFIVKKSDEYTFATSDAHNTMNVDD	122
Db	330	TSSTLENDSL--ELDHGNSNDSIDLKAYNIDIEL--HTYST--QILKATLDN	377
OY	123	-----QEVINKAS-SKKITLEKRLYQI-KIO	148
Db	378	IQKLGDCNDLVKDOCKELRELSTALYDLKIQITSVINRENDISNNIDIVSKNLMEIDAIQ	437
OY	149	YORE-----NPTKGLFPLKYWTDSQKKE	173
Db	438	YNFEYKEIPIPNVEYKTLDDTKNAYIYKKAELKANDINKTRKEDDIIYFNDDLELEKSL	497
OY	174	VISSDNIQLPELKOKSSNRKRKSTSAQPIVPPDRDNG--IP--DSLEWEGYTVD---	224

Db	498	TLSSNEHEITTYONSTNS-----FSQIKNNIDIDKEMKTLIPMLDELLNECHNIDSLY	553
Qy	225	---VKNRTELSPWISNIEHKGLT---KYKSSPEKMWSTAS---DPYSDPEKVTGRID	273
Db	554	NFIIRNIOIGINDIKNIREQENDTICEYEOIONNNFIKSDISIFNKYDHIKVDYIS	613
Qy	274	KVNSPEARHPPLVAAYPVHVD--MENI-----ILSKND-----OSTO-----NTDSE	314
Db	614	NNIDVAKHSHLSSEHYVNTNIIENIMYSIYEINEDTMSLSEETQDLLELYENKE	673
Qy	315	TRTISKNTFSR--THSEVHGAENVHANTSTKTHTSEVHG-----NAEV	358
Db	674	KNIINNNKYVHFENKLEIENSLETYSNSTENKINETQNDIDLKNEFNNIKTKINDKV	733
Qy	359	-HAVAIDHSIAGEPMAETWG--LNPADARL-----NANIRVYNGTAPIYVLPPTT	410
Db	734	KELVHVSTLTLESTIOYFNNLYGDLSMSNIDYKYKDIYDINVELKAKVLYENITNULGRI	793
Qy	411	SLVLGKNOGLATTIKAKENQLS---QILAPNNYPSKNLAPIALNADDFSSPTITMANYQ	467
Db	794	NTFI---KELDKYODENNGIDKYEIENKENSYIIL--LKEKANNNKENFS-----K	840
Qy	468	FLELEKTKOLRLDIDYGYIATYFNENGRVDRDGSMMNSEVLPQIOETARIIFENCK	527
Db	841	LLOIKIKNETEL-----YNNN--IKDD-----IMTKGSV	869
Qy	528	NLVERRIAAVNPDDPLETTKPDMTLKEALKIAGFENEPGNLOYOGKDIPEDFENFDOT	587
Db	870	NNIKQFSS-----NLPLKEKL---FQWEMMLNI-----	896
Qy	588	SQNKKNQDLAELNATNITYVL-----DKIKLAKANNILIRPKRF--HYDRNNIAVGADES	639
Db	897	-NNIMNETKRISNDVYNTTTDODIENNKNKNNNNIETIDKLIDHIKIHNEKIQAETL	955
Qy	640	VKVEAREHVINNSTEGLLIDDKIRKILSGYIV-----IEDTEGLKE--VINDRYV--	690
Db	956	IIDDAKRRV-----KEITDINKAFMEITENYNNENGVIKSAKNNIYDEATYLNELDKF	1010
Qy	691	MLNITSLSRODGKTFIDPKKYNDK	713
Db	1011	LTKLNELLSHNNN--DIKDLGDE	1031

RESULT	11	
Q9BK45		
ID	Q9BK45	PRELIMINARY; PRT; 3254 AA.
AC	Q9BK45;	
DT	01-JUN-2001 (TrEMBLrel. 17, Created)	
DT	01-JUN-2001 (TrEMBLrel. 17, last sequence update)	
DT	01-JUN-2001 (TrEMBLrel. 17, last annotation update)	
DE	RETICULOCYTE BINDING PROTEIN 2 HOMOLOG B.	
OS	Plasmodium falciparum.	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX	NCBI_TaxID:5833;	
RI	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE-21101060; PubMed-11160005;	
RA	Trigila T., Thompson J., Caruana S.R., Delorenzi M., Speed T.,	
RA	Cowan A.F.;	
RT	"Identification of Proteins from Plasmodium falciparum That Are	
RT	Homologous to Reticulocyte Binding Proteins in Plasmodium vivax."	
RL	Infect. Immun. 69:1084-1092(2001).	
DR	EMBL: AF312917; AAK19245.1; -	
SO	SEQUENCE	3254 AA; 382876 MW; 6F9CAFA5A6167BA CRC64;

	Query Match	5.28;	Score 194.5;	DB 5;	Length 3254;
	Best Local Similarity	19.68;	Pred. No. 0.017;		
	Matches 169;	Conservative 135;	Mismatches 290;	Indels 269;	Gaps 41;
QY	5 KVLIDLMALSTLVASTGKLELYIQAEVKQENMLLNESSESSGGGLGYTFSDNLFPQAPMNV	64			
	: : : : : :	:	:	:	:

QY 5 KVLIPMALSTLIVSSGTNLEVIQAIEVKQENRRLNSESSESSQGLGYYFSDNLFOAPMV 64

Db	1920	KHVILAFSGIIYMSDPTNFEIRPENDELNDLN-----	LDLYFERKHEI	1965
Qy	65	TSSTTGDLSTSSSELENI--PSEQIFQASIMSGFIVKXKSDSEYIPASDNIHTMAYD	122	
Db	1966	TSTLENSDL---ELDHLGNSDSEIDNLKYYDIEL-----HYST---QILKYLDN	20133	
Qy	123	-----	DEVINKASN--SNKIRLEKGRLOYI-KIQ	148
Db	2014	IQKLKGDGNDLYKCKRELRELSTALVDLKIQITSVINRENDSSNNIDIVSKNLEITAIQ	20733	
Qy	149	YQRE-----	NPEKGLDFKLWTSQNKKE	173
Db	2074	YNFEYKEIFEDNVEEYKTLDDTFKAAIVYKKAELIKHWIDINKREDIDYFNDLDELEKSL	21333	
Qy	174	VISSNNOLPRELKOKSSNSKKRSTAGRPVPRDNG---IP--DSEVEGYVD-----	224	
Db	2134	TLSSEMEIKITVONSTNS---PSDINKNIDIDEMKTLIPMLDELINBGNHIDISLY	21899	
Qy	225	--VKNKRTFLSPWISNIHKKGLT-----KYKSSPKWSTAS---DPYSDFEKYGTID	273	
Db	2190	NFIIRNIQIKGNDIKNIRBENDNTNCFEYIGNNYPFKSDISJFNKYDHIKVDNIS	2249	
Qy	274	KNVSEARHPVLAYPVYHD--MENI---ILSKND-----OSTQ-----NDSE	314	
Db	2250	NNIDVYNNHNSLSEHVIINNTIENMTSIVEINDTEMSNLEETODKLELYENFKKE	23099	
Qy	315	TRTISKNTSTSR--THTESEVGNAEVHANTSTRTSHSEVHG-----NAEV	358	
Db	2310	KNIIINNKKYIVHFKLKEIENSLETYSNSTENKINETQNIDILKNEFNKIKTKINDXV	23699	
Qy	359	HAAVAIDHSLSLAGEKRTWAETAG---LNTADTARL---NANIRYVNTGTAPIYVNLPTT	410	
Db	2370	KELVHVDSTLLESIOGFNNLYGDLSMSNIQDYKVEDINNVLEKRVKLYIENTITNLGRI	24299	
Qy	411	SLVGLKNOTLATIYAKENQNS---QILAPNNYPSKNLAPILNKODPFSSPTIMANNQ	467	
Db	2430	NTFI---KELDKYODEMNGDKDKYIEINKENSTIYK-LAEKANNKLENFS-----K	2476	
Qy	468	FLELEKTKQLRLDQYVGNMIATYNFENGVRVDTGSSNMSEVLPOLIOETARIIFNGDL	527	
Db	2477	LLOINKRETEL-----YNINN--IKDD-----IMNTGKSV	25055	
Qy	528	NLVEERIAAVNPSDPLETTKPDMLTKEALKIAGFWEPCGNIQYOGKIDTEPDFNFQOT	587	
Db	2506	NNIKQKSS-----NLPLKEKL--FQMEEMLNT-----	25322	
Qy	588	SONIKNOLAELNATNIYTVL-----DKIKLAKANNILIRDKRF--HYDRNNIAGADES	639	
Db	2533	NNIINNETKRISNTDAYTITLQDIENNKNNKNNNNIETIDKLIDHIKIHNEKIQAIL	2591	
Qy	640	VYKAEHREVYNSSREGLLIWDIKIKLSGIYV-----IEPTBGLKE---VINDRYD--	690	
Db	2592	IIDDAKRV-----KETTDINAKAFMEITEYNNENNGVYKAKNIYDKATYVNLNELDKF	2646	
Qy	691	MLNISLSLRDGGKTFIDFKKKNKD	713	
Db	2647	LKLINELSHNNN--DIKIDGDE	2667	
RESULT	12			
Q9BK46				
Q9BK46		PRELIMINARY;	PRT; 3130 AA.	
AC	Q9BK46;			
DT	01-JUN-2001 (Tremblrel. 17, Created)			
DT	01-JUN-2001 (Tremblrel. 17, Last sequence update)			
DT	01-JUN-2001 (Tremblrel. 17, Last annotation update)			
DE	RETRICULOCYTE BINDING PROTEIN 2 HOMOLOG A.			
OS	Plasmodium falciparum.			
CC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=5833;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21101060; PubMed=11160005;			

RA Triglia T., Thompson J., Caruana S.R., DeIorenzi M., Speed T.,
RA Cowman A.F.:
RT "Identification of Proteins from Plasmodium falciparum That Are
RT Homologous to Retroviral Binding Proteins in Plasmodium vivax".
RL Infect. Immun. 69:1084-1092(2001).
EMBL: AF121816; AFK19244.1; --
SO SEQUENCE 3130 AA; 370413 MW; 13D973DB89DB2026 CRC64;

Query Match	5.18;	Score 191.5;	DB 5;	Length 3130;
Best Local Similarity	19.68;	Pred. No. 0.025;		
Matches 169;	Conservative 134;	Mismatches 291;	Indels 269;	Gaps 41;

Qy	5	KVLPLMALSTIIIVSYSGNLEVLQAEVKQDRELLNSESSESSOGGLGYFSDLNFOAPMV	64
Db	1920	KHVTKLANFSGIYMSOTNTEIPRENDENLNL-----LOLYFERKHEI	1965
Qy	65	TSSTGTGDISSELENI--PSENOFQSAIWSGFIKVKKSDYEFPATSADNHVMMVDD	122
Db	1966	TSTLENDSDL---ELOHLSGNSDESIDNLKVYDIIEI-----HYTST---QILAKYIDN	2013
Qy	123	-----OEYINKASN-SNKIRLEKGLYOI-KIO	148
Db	2014	IQKLKGDONDLVKDCKELRELBALYDLKIQIYVSINRENDISNNIDIYSNKLNEIDAIQ	2073
Qy	149	YQRE-----NPREKGLDFLYWTDSQNKKE	173
Db	2074	YNFEKYKEIDNVEEYKTLDDTKNAYIVKKAELKAVDINKRKEDLDIYFNOLDLEKSL	2133
Qy	174	VISSDNIQBPCLKOKSSSKKRSAGTPVPRDNDG---IP--DSLEVEGYVD---	224
Db	2134	TLSSNEHEIKIYVONSYSN-----FSDINKNINDIKEMKTLIPMLDELLNEGHNNISLY	2189
Qy	225	---YKNRTFELPMISINHEKKGLT-----KYSPEKMTAS-----DPVSEKXTGRI	273
Db	2190	NFIIRNIQIKINGDKIRQENDNTICEFYIONNNTFKSDISFPNKIDYHIKVDNTS	2249
Qy	274	KVNSPEARHPLVAAPRYVHD--MENI--ILSKNED-----QSTO-----NTDSE	314
Db	2250	NNIDVKNHNSLSEHYINATNIENIMTSIYEINDETMNSLDETQDKLELYENFKKE	2309
Qy	315	TRTISKMTNSR--THSEVGNABEVIANSTSKTHTSEVHG-----NAEV	358
Db	2310	KNILNNNKKIYVHFNKLEIENSLSETYSNISTENFKINETQNDILIKNEFNNIKTKINKY	2365
Qy	359	HAYVIDHSISLAGEPRTAETMG---LNTADTARL-----NANIRYVNTGTAPIVAVLPPT	410
Db	2370	KEIYHVDSTLIESIQFNNLYODLMSNIQDYKYEIDINVELKAKVLYIEMITMLOGRI	2429
Qy	411	SLVIGKQOTLATIKAKENQUS---QILAPNNTYPSKNLAPILAMODESSPTITMANTQ	467
Db	2430	NTFF--KELDKYQDENNGIDKYEIENKENNSYIK-LKEKANNKENNS-----K	2476
Qy	468	FLELEKTKOLRLTDQYUAGIATYFNEENGVRVDGTSNMSEVLPQIOETARIIFNGDL	527
Db	2477	LLOIKRKNHEI-----YNNNN--IKDD-----IMATGCSV	2505
Qy	528	NLVERIAAVNPSDPLETTKPDMTLKBAKIAFGNEPNGNLOYGOKDITTEDFNFDQOT	587
Db	2506	NNIKOKFSS-----NLPLEKELT---FQMEEMELNI-----	2532
Qy	588	SONKNOLOAELNATNYTYVL-----DKIKLAKKNILIRPKRF--HYDRNNIAYGADES	639
Db	2533	NNNNMETKRISNTAAATNTTLODIENKKKENNNNNIETIKLDIHIKINEKIQAEIL	2591
Qy	640	VKREAREVINSSTEGLLIDDKIRKISGYVE-----IEDTEGLKE---VINDRYD--	690
Db	2592	IIDDAKKKV-----KETDINIKAFNEIYENVNNENGVYKSAKNIVDEATYALNNEIDKF	2646
Qy	691	MLNISTSLRQDGKTFIDPKKYNDK	713
Db	2647	LKLNELLSHNNN--DIKDGDGE	2667

RESULT 13
ID 09P008 PRELIMINARY: PRT; 4688 AA.
AC 09P008;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, last annotation update)
DE HYPOTHETICAL PROTEIN U0482.
GN U0482.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
ON NCBI_TaxID=134821;
RX SQUENCE FROM N.A.
RC STRAIN=SEVOVAR 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Leikowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
urealyticum.";
RL Nature 407:757-762(2000).
DR EMBL: AE002145; AAF30894.1; -
DR InterPro: IPR001152; Thymosin_b4.
DR InterPro: IPR000626; Ubiquitin.
DR Pfam: PF00240; ubiquitin; 1.
DR SMART: SM00152; THY; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 4688 AA; 534880 MW; B53ABFAFEE1997E CRC64;

Query Match 5.1%; Score 190.5; DB 2; Length 4688;
Best Local Similarity 20.7%; Pred. No. 0.05;
Matches 168; Conservative 132; Mismatches 341; Indels 171; Gaps 39;

QY 18 VSTGNGLEVI-QAEVKOENRLNSESSESSOGLGYYFSDLFQAPMYVTSSTGDLSPS 76
Db 3676 INKTGKQEVFVSKYGLKSNQLYK-----LVQVYILD-NIHONIDETKRIEFDHNV-S 3725
QY 77 SELENPSENOYFOSAIWSGFIKKKSDYTFATSDAHVMTMVDQEVINKASNSKIR 136
Db 3726 KELEINGVTMISGHGMKSPDTTANFEKIEFQ-----DDNDVLINIDATYKFK 3776
QY 137 LEKGRLYQIKIYORENPTKGLFKLYMDSQKKEVSSDNLQLEIKQSSN-----S 192
Db 3777 DEHNINIKQKIVRIKEN-----NDMLIKGI--DNLN-PEYKYLEINIELSK 3820
QY 193 RKRSTSGPTVPDRD-----NDGIPDSLEVEGYTVYDKKRTFLS 233
Db 3821 PLKTHNLASINDKENISLITETGNVFLVYIQONDITNDTOQTIVTILSGVNSK-YNG 3879
QY 234 PWISNIHEKKGLTKYKSSPEKMWSTADPYSDFEKVTGRIDKNVSPEARHPLVAAPYIVH 293
Db 3880 RQIKVYKKDNNNVYESS--LITLQKNDYQLLSNLNSN---REYFEKEIEINHI 3921
QY 294 -----DMF-----NIIISKNEOSTQNTDSE-----TRTISKNTSTSTHSEVHG 335
Db 3932 SNTNNEDEKLKNGVSTFTQTKNTTVQNMDSATIVGRGVNFNKI-KSEDKILENN 3990
QY 336 AEV---HANTSTRTHTSEVHGNAEVAVALIDHSLSLAGERTMAETGLNT---ADTARL 389
Db 3991 QQVVAFAPEKTIKDTITWLOQYTRPLKDVTSDFK-----EGTNAHDLSSNVEKEETTYK 4045
QY 390 NANIRVYNTGTAIYVNLPTTSLVGLKNQTLA-----TKAKENQLSOLLAPNNYPSK 443
Db 4046 LVRIQFNKPTKAKNNINSENINVLIDNTNSINSNYEFTTKVGDHKLINITSSNNVWNS 4105
QY 444 NLAPIALN-AQDESSPTITMNY--NQLFELEKTKQLRLTDQVYGYNIATINFEKNGR--- 497
Db 4106 QTINFTLSGVKKSMGKRIKILSKYKSDNTSEIHTEWELISNKQYVILNLNLRKNTYT 4165
QY 498 ---VRVDYGSNMSEVLQIQTETARLIFNGKDLVLVERIAAIVAPSPDLETT-----K 547

Db 4166 LIDVKLIDNNNVSDPEKGNLTNSFITRTSAIVNLNIEISNRASFTLKSITIKINLD 4225
QY 548 PDMTLKEA-----LKIAFGNEPENGNIQYQKQITTEDFNFDOOTSONIK-NOLAE 597
Db 4226 PDVNLRRKQDAITYGNKKQAMGFIVSGNIKYLTAFLVDLNFN-DKVNINISFNKPS 4284
QY 598 LNTATITVLDKIKLNAKMNILINDKRFHYD-----RNNIYAGA--DESVYKAHRE 647
Db 4285 IAAEN-----IGIDKSNNTI-----YNDSTIKLELNNDIYNGPINKETIYKMANOK 4332
QY 648 VINSSTGBLNLINDKIRKILSGYVEIEDETEGLEKVINLND--RYDMLNSSL-RQDKTE 704
Db 4333 --NNIDVDLGIQINPKTAHNLN-FLAKFKSTN-----NDIETINVIKSSLVNNDGRTS 4383
QY 705 IDEKKYNDKLEPLIYSNPNYK-VNYYAYATKENT 735
Db 4384 IRTFLNMLK-----ANKLYSLVDVYVYLVNNS 4410

RESULT 14
ID 09CF64 PRELIMINARY: PRT; 1072 AA.
AC 09CF64;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, last annotation update)
DE UNKNOWN PROTEIN.
GN YQFG.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
ON NCBI_TaxID=1360;
RX SQUENCE FROM N.A.
RC STRAIN=IL1403;
RA Bolotin A., Wincker P., Manger S., Jallion O., Malarne K.,
RA Weisenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis.";
RL Genome Res. 0:0-0(2001).
DR EMBL: AE006392; AAK05715.1; -
KW Complete proteome.
SQ SEQUENCE 1072 AA; 113056 MW; 464446E2656CAA08 CRC64;

Query Match 5.0%; Score 187; DB 2; Length 1072;
Best Local Similarity 18.8%; Pred. No. 0.01;
Matches 146; Conservative 151; Mismatches 333; Indels 148; Gaps 29;

QY 7 LIPALSTIIVSSTGNGLEVIQAEVKOENRLNSESSESSOGLGYYFSDLFQAPMYVTS 66
Db 319 LYPISSEASVYDNTLNTSSISLSSISS-----SQTENSQSG-----A 355
QY 67 STGDLSPSSSELENPSENOYFOSAIWSGFIKKKSDYTFATSDAHVMTMVDQEV 126
Db 356 STAEISTDSSENSSLSSNO-----INSNSNSEKDSQSSLGSSMSSESEHSNSNTI 410
QY 127 NKASNSKIR-----LEKGRLYQIKIYORENPTKGLFKL-----YWTDSQNKKEVYS 177
Db 411 NETNNSSEITNIIPLPSNPTESNVSDDQSSSEASINSNSISLSPSNISSTSDSATSATSS 470
QY 178 DNLQPLKQKSSNSRKRSTSGPTVPDRNDGI-----PDSLEVEGYTVYDKKRTFLS 233
Db 471 DFSVVAEYANNISLAVNNSSSVLSSTADNLGINOSGDNLTKD--SSEISTSGAFLS 528
QY 234 PWISNIHEKKGLTKYKSS--SPEKWS-----TASDPYSDFEKVTGRIDKNVSPPEAR 281
Db 529 ---SNQTSSEKSTNSNSISLSPSNISSTSVLESTSSNFSNAEVAANNSLASVNNSS 585
QY 282 HPLVAAPYIVAVDM-----ENIILSKNEQ-----STQNTDSEFTRTISKNTSTSRPHTS 330
Db 586 SVLSSTSTADNLEINQGSQNLTKDSSSEISTSGAFLSSNQTSSEASNSMSINSPLSL 645

OY 331 EVHGNAEVAHANTSTRTHTSEVHGNAEVAHAVIDHSLSLAGERTMAETMGLNTADTARLN 390
DB 646 SLTNSSEATNGSNS-SEATKYDNSTSHSNLNS-----GSDSDSDSDSDSDSNL- 699
OY 391 ANIRVYNTGTADLYNVLPTSLVLGKNOTLATIKAKENOLSO-----ILAPNNYPSKNLA 446
DB 700 -----SSPRLETNOTISSKPSFEVNNIISEPKVSSNSNSQOENSTDH 741
OY 447 PLALNAODEFSSTPTTMANYNOFLEKTKQLRLDIDQYGNATYV--FENGRAVVDTG 503
DB 742 EMTNPKSSISTSPITSSSQOKESQSN--LLNTTEGINNPTFENSSNSAAIL- 797
OY 504 SMMSEVLPQIOETARILF-----NGKDL-----NLVERIAAVNSDPLETTKPKM 550
DB 798 -SYSNNSSSESTGCLYISNEAORDGSEISHSLPSSNSNENNVSSIQOAILSESSKST 856
OY 551 TLK-EALKIAGFNEP-NGNLOYOGKIDTEPFDFNPDQSONIKNOELATNTIYTD 608
DB 857 MKRSSLSIINSTSHQNDNOSNSD-----EVKSNNVESILGOLNSISNKTTHN 908
OY 609 KI--KLNAKMILIRDKRPHYDRNNIAVGADESVYKAEHREVINSSTEGLLNLDKDIR 665
DB 909 SLTQSLAVIYTLPSKSKVTNEKENSNTVSEKLIKTPQKN-DESONLGQITALDSLTFN 967
OY 666 KILSGIYIEDETEGLKEYIN-----DRYDMLNISLRD-GTTFIDFKKYNDKL 714
DB 968 K-----EVTMEDSKTPVDRVLNENGRSONNKTSTIARDKNKVFPRSEFNSKI 1018

RESULT 15
O46149
ID O46149 PRELIMINARY; PRT: 2178 AA.
AC O46149; O46147; O46148; 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE ALPHA-TOXIN.
OS Clostridium novyi.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC19402;
RX MEDLINE=95342160; PubMed=7616958;
RA Hofmann F., Hermann A., Habermann E., von Eichel-Streiber C.;
RT "Sequencing and analysis of the gene encoding the alpha-toxin of
RT Clostridium novyi proves its homology to toxins A and B of Clostridium
RT difficile.";
RL Mol. Gen. Genet. 247:670-679(1995).
RN [2]
RP SEQUENCE OF 1204-2178 FROM N.A.
RC STRAIN-ATCC19402;
RA Hofmann F., Habermann E., von Eichel-Streiber C.;
RL Submitted (JUL-1993) to the EMBL/Genbank/DBJ databases.
DR EMBL; Z48636; CA88565.1; -
DR EMBL; Z23280; CA80818.1; -
DR EMBL; Z23281; CA80819.1; -
DR InterPro: IPR002479; CW-binding.
DR Pfam: PF01473; CW-binding.1; 10.
SQ SEQUENCE 2178 AA; 250134 MW; 980ADCE031CA75A CRC64;

Query Match 4.9%; Score 183.5; DB 2; Length 2178;
Best Local Similarity 20.6%; Pred. No. 0.043;
Matches 152; Conservative 109; Mismatches 241; Indels 235; Gaps 34;

OY 122 DOEVIKASNSKIRLEKRLYQI-KIYOAREN-----PTEKGLDFKL---- 163
DB 202 DDIINFLSNYK-----YDIGKLNQKNNNNKMAIGAATDINTENILTNKLKSYX 253
OY 164 YWTDSONKKEVISSDNLQLELKQSSNSKRKRSAGTTPVDRNDGIDPSLE-VEGTY 222

DB 254 YDELLITNNLAAASDILRLAILKKY-----GYVCDIDFLUPGN 292
OY 223 VVKNRRTFLSPWISN-----IHEKGLTKYKSSPEKWTASDSEFEKTYGRI 272
DB 293 LSLFNDISKPNMGDSYWEAIFEAIAINEKTKLMNN-----PYKMEQVPSLI 340
OY 273 DKNVSEAR-----HPLVAAPVIVHDMENILSKNEDOSTONTDETTRISKNTS-TSRT 327
DB 341 KERILSFVNHHIDINDILPLGDIKISQLEILRLKAAGKRTFSNAFIISNDSLTLLNN 400
OY 328 HTSEVHG-----NAEVAHANTSTRTHTSEVHGNAEVAHAVIDHSLSLAGERTMAETMGL 381
DB 401 LISQLENREILNLSITIOEKFKICETDYSTINSVSELVLETTTPKRLNSMDGSSFFQOIIIGYL 460
OY 382 NTADTARLANIRY-----VNTGTAPIYVVLPTT-SLVGKNOTLATIKAKENOLSOILA 435
DB 461 SSGFKEVNSYTFPSPGNITSSATCGTYHFIKNTFPMLSQODET-----FEA 508
OY 436 PNNYPSKNLAPALNAODEFSSTPI--TMANYNOFLEKTKQLRLDIDQYGNATYV 492
DB 509 SNMLYFSK-----THDEFKSSWLLRSNIAEKEFOKLITYIGR-----TLN 549
OY 493 FENGR-----VAVDGSNNSEVLPQIOETARILFNGKDLNVERIAAVNSDPLETTK 547
DB 550 YEDGLNFKWKRYTT-----SELKVIEVNSKTYIENYDLNMI-----LQIOG 593
OY 548 PMTLKALKIAGFNEPNGNLOYOGKIDTEPFDFNPDQ--QTSONIKNOELN----- 599
DB 594 DDISYSAVNV-FGKN-PKKSILIQGVDDPFAVTFYENCIYVSDNINNLISRFNDKKIK 651
OY 600 -----ATNIYVLDKIKLN-----AKMNL----- 619
DB 652 LVLIGHGVNFPKLFEGKTVNDLYTNITKPKLOHLREGVILKKKYLIKINILGCMFT 711
OY 620 -----TRDKRFH-----YDRNNIAVGADESVY--KEHREVINSSTEGILL 658
DB 712 PYVDINSTVFGKLFNKISDLOPKGFSKNQLEISAKTAIRINREKREVLDFGK-WVS 770
OY 659 NIDKDIRKILSGIYV--EIDTEGLKEYINDRYDMLNISLRDQKTFIDFKKYNDKLP 715
DB 771 NMDLIAEQISNKKYVYVWNEVENT-----LSARVEQLN-----KAAEFKAKDINSIT 815
OY 716 LYISNPYK--VNYA 729
DB 816 QTTNMOELKOSLVNTYA 832

Search completed: December 2, 2001, 13:55:15
Job time: 550 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:52:01 ; Search time 44.8 seconds
(without alignments)
601.532 Million cell updates/sec

Title: US-09-747-521-4_COPY_1_735
Perfect score: 3765
Sequence: 1 MKRRKVLPLMALSTILVS.....LYISNPXKVAVATKENT 735

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3652	97.0	764	1	PAG_BACAN
2	211.5	5.6	192	1	P13423 bacillus an
3	183	4.9	1803	1	P47024 saccharomyc
4	177	4.7	1658	1	Y667_YEAST
5	172	4.6	1271	1	P47880 mycoplasma
6	171	4.5	1230	1	SMC3_YEAST
7	166.5	4.4	1928	1	P08964 saccharomyc
8	160	4.2	1790	1	P25386 saccharomyc
9	159	4.2	2116	1	P08799 dictyostell
10	158.5	4.2	1487	1	P53094 saccharomyc
11	158	4.2	1885	1	O13735 schizosacch
12	156.5	4.2	803	1	P09959 saccharomyc
13	155.5	4.1	1018	1	P14738 staphylococ
14	155.5	4.1	1541	1	P42782 haemophilus
15	155.5	4.1	1545	1	P45385 haemophilus
16	155	4.1	1630	1	P04932 plasmodium
17	155	4.1	1639	1	P04933 plasmodium
18	154.5	4.1	1000	1	P03512 saccharomyc
19	154	4.1	1251	1	P00799 plasmodium
20	153	4.1	1570	1	P54573 dictyostell
21	153	4.1	1744	1	O01550 xenopus lae
22	151.5	4.0	2334	1	O07933 bacillus su
23	151.5	4.0	2869	1	O00798 plasmodium
24	151	4.0	1957	1	O10411 schizosacch
25	150.5	4.0	1208	1	P09251 plasmodium
26	150	4.0	1726	1	P50495 plasmodium
27	149.5	4.0	1114	1	O12749 saccharomyc
28	149.5	4.0	1420	1	P38931 saccharomyc
29	149	4.0	1577	1	P16466 proteus mir
30	148.5	3.9	979	1	P115 MYCHR
31	148.5	3.9	1233	1	P43597 mycoplasma
32	148	3.9	1024	1	P97434 mus musculu
33	146.5	3.9	1165	1	P53950 saccharomyc

34	146	3.9	730	1	GLN3_YEAST	P18494 saccharomyc
35	146	3.9	770	1	ACE2_YEAST	P21192 saccharomyc
36	146	3.9	3418	1	BRC2_HUMAN	P51587 homo sapien
37	145.5	3.9	1093	1	SWI4_YEAST	P25302 saccharomyc
38	145.5	3.9	1230	1	ST20_CANAL	O92212 candida alb
39	145	3.9	635	1	PIB2_YEAST	P53191 saccharomyc
40	145	3.9	976	1	SCPI_HUMAN	O15431 homo sapien
41	145	3.9	1358	1	SIR4_YEAST	P11978 saccharomyc
42	145	3.9	1664	1	INT1_CANAL	P53705 candida alb
43	145	3.9	1726	1	MSPI_PLAFC	P04934 plasmodium
44	144.5	3.8	633	1	IPAA_SHIFT	P18010 shigella fl
45	144.5	3.8	678	1	YNC7_YEAST	P53968 saccharomyc

ALIGNMENTS

RESULT	1	STANDARD;	PRT;	764 AA.
PAG_BACAN				
ID	P13423;			
AC	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	PROTECTIVE ANTIGEN PRECURSOR (PA) [CONTAINS: PA-20; PA-63].			
GN	PAG.			
OS	Bacillus anthracis.			
OC	Plasmid PXO1.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1392;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=89172073; PubMed=3148491;			
RA	Welkos S.L., Lowe J.R., Eden-McCuthan F., Vodkin M., Leppla S.H.,			
RA	Schmidt J.J.;			
RT	"Sequence and analysis of the DNA encoding protective antigen of			
RT	Bacillus anthracis.";			
RL	Gene 69:287-300(1988).			
RN	[2]			
RP	DOMAINS.			
RX	MEDLINE=91332080; PubMed=1651334;			
RA	Singh Y., Kimpel K.R., Quinn C.P., Chaudhary V.K., Leppla S.H.;			
RT	"The carboxyl-terminal end of protective antigen is required for			
RT	receptor binding and anthrax toxin activity.";			
RL	J. Biol. Chem. 266:15493-15497(1991).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).			
RX	MEDLINE=97192099; PubMed=9039918;			
RA	Petosa C., Collier R.J., Kimpel K.R., Leppla S.H., Liddington R.C.;			
RL	"Crystal structure of the anthrax toxin protective antigen.";			
RL	Nature 385:833-838(1997).			
CC	-1- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,			
CC	AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE			
CC	DEATH. PA IS THOUGHT TO BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC			
CC	CELLS, THEREBY FACILITATING THE INTERNALIZATION OF LP OR EF. PA			
CC	ASSOCIATED WITH LP CAUSES DEATH WHEN INJECTED, PA ASSOCIATED WITH			
CC	EF PRODUCES EDEMA. PA INDUCES IMMUNITY TO INFECTION WITH ANTHRAX.			
CC	-1- SUBUNIT: ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT PROTEINS, A			
CC	PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN EDEMA FACTOR			
CC	(EF). NONE OF THESE IS TOXIC BY ITSELF. ONCE ACTIVATED, PA FORMS			
CC	HEPTAMERS WHICH INSERT INTO MEMBRANES AND FORM CATION-SELECTIVE			
CC	CHANNELS.			
CC	-1- SUBCELLULAR LOCATION: SECRETED.			
CC	-1- DOMAIN: THE C-TERMINAL PART OF PA IS REQUIRED FOR RECEPTOR BINDING			
CC	AND TOXIC ACTIVITY.			
CC	-1- PTM: PROTEOLYTIC ACTIVATION BY FUZZIN CLEAVES THE PROTEIN INTO TWO			
CC	PARIS, PA-20 AND PA-63, THE LATER HEPTAMERIZE.			
CC	-1- SIMILARITY: TO C. PERRINGENS IOTA-B TOXIN AND TO VIP1 TOXINS IN			
CC	BACILLUS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			

CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 ON NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1679;
 RX MEDLINE=97103775; PubMed=8948101;
 RA Cielepuch C., Kordes E., Pujol A., Jauniaux J.-C.;
 RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
 reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,
 SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, CCF7, GZF3, two tRNA genes,
 RT three remnant delta elements and a Ty4 transposon."
 RL Yeast 12:1471-1474(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z49389; CA89409.1; -;
 DR SCD: S0003649; J01113W;
 DR InterPro: IPR001584; Rve.
 DR InterPro: IPR001878; ZnF_CCHC.
 DR Pfam: PF00665; rve; 1.
 DR SMART: SM00343; ZnF_C2HC; 1.
 KW Transposable element; Hypothetical protein;
 SQ SEQUENCE 1803 AA; 207693 MW; 16DCD7284A8D52D3 CRC64;

Query Match 4.9%; Score 183; DB 1; Length 1803;
 Best Local Similarity 20.1%; Pred. No. 0.1;
 Matches 167; Conservative 117; Mismatches 283; Indels 264; Gaps 40;

QY 25 EVIQAQVQENRLNLESESSQGLGYFSDLNQAPVWYTSSTGDISPSSSELENIPS 84
 DB 1073 ELIOEOLKKT---HETSFREGSIG---TWKFRNTNEISLKTGDTSLPTKTESINN 1126
 QY 85 ENQVFOAINGCFIKVKKSDDEYFATISADNHTVMVDQEVINKASNKRLRLEGRILQ 144
 DB 1127 HH-----SNLY-----STNKE----- 1138
 QY 145 IKIQOYQREN---PTKGLDFKLYWTDSONKKEYISSDNQLPELKQSSRKKESTSA 200
 DB 1139 ---KFEKENHHPRPIEDIVDM---SDQTMESNCQDGNLKLKLYTDKAVPDINGINV 1190
 QY 201 GPVYPRDNDGIPDSLEVEGYTVYKKNKRTFLSPWISNIHEKK-----GLTKY--- 248
 DB 1191 SPR-----LEQNIASGSPVQTVNKSAPLNKFSLSLANKRRKRHRDKNNSLTSELE 1242
 QY 249 -----KSSPEKMTASDPYSDPEKYTGRIKRVNSEAKHPLVAAPYHAVDMEN 297
 DB 1243 RDKRKRKKNRYKLLPDDNNEYSAKIRAIYYNEAISKNPDLEKHEKQAY---HKEION 1299
 QY 298 IILSKNEDQSTQNTDSEYRTISKN--TSTSRTHTSEVHG--NAEYHANTSTSRHTSEVH 353
 DB 1300 LKDMKVPDVYKYSRSE---IPDMLYPTNTITFKKRGITKARIVCGDTPQSPDTYSVI 1356
 QY 354 GNAEYHVAVIDHSLAGERT-MAETMGLNTADT-ARLNANIRYVNTGTABIVNLP--TT 410
 DB 1357 TTESLNHNHIKIFLMIANNRMFMKTLIDINHAFLYAKLEED-----IYIPHPDR 1406
 QY 411 SLVIGKQGLATITAKENQ-----LSQI-LAPNNYTP-----SKNLAPIALNADD 455
 DB 1407 RCVVYKLLKALYGLKQSPREKMDHLROYLNGIGLKDNSTPGLYOTEDKNM-LIAYVDDC 1465
 QY 456 FSSPTIMYNYQFL-ELEKTKQLR-----LDTDOVYENINATYNNENGRV----- 498
 DB 1466 VIASNQRORDEFINKLSNELKITGGLIDVDLTD--ILGMDLYNNKRLTIDITLTKSF 1524
 QY 499 --RVDGTSNMSEVLPOIQTETARIIFNGKDLNLYERRIAAVNP--SDPLETTKPD-----M 550
 DB 1524

DB 1525 INRMD--KKNYNEELKIRKRSIPHMSTYK-----IDPKKDVLYOMSEEFROGVL 1571
 QY 551 TLKEALKIAPGFNENPNNGMLQYQGXDTFEDFNPDQOSTQNKQLAEALNATNIVYVLKI 610
 DB 1572 KLQQLL-----GELNTV-RHKCRYDIEFAVKKARLVNPHEVFYMTIKIOLY 1620
 QY 611 KLNKAMNILIRDKRREHYRN-----NIAVGADSEYVKAHRE---VI----- 649
 DB 1621 -----VRYDIGHYRDCNKKKVAITIDASGSEYDAQSRIGVILWGMNIFNYS 1673
 QY 650 NSSTEGLLNIDKDIRKILSGY-----IYEIEDT---EGLK--- 682
 DB 1674 NKSTNRVYSTAEALHAYEGVADSETLVTLKELGEGDNNDIYMTDSIPAQGLNSRY 1733
 QY 683 -----EYINDRYDMINISLQDGKTFI-----DRKKY 710
 DB 1734 QQPKREFTWIKTEIIEKIKESIKLKITGKGNIADLLTRPVASDEKRF 1784

RESULT 4

YM67_YEAST STANDARD; PRT; 1658 AA.
 ID YM67_YEAST
 AC Q03661; Q04988;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHEETICAL 187.1 KDA PROTEIN IN GUAI-ERG8 INTERGENIC REGION.
 GN YM6219W OR YH8261.13 OR YH9959.01.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE OF 1-711 FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
 RA Walsh S.V.;
 RL submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 608-1648 FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z49809; CA89934.1; -;
 DR EMBL: Z49839; CA990190.1; -;
 DR SCD: S0004832; YMR219W.
 KW Hypothetical protein;
 SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;

Query Match 4.7%; Score 177; DB 1; Length 1658;
 Best Local Similarity 19.3%; Pred. No. 0.18;
 Matches 159; Conservative 132; Mismatches 300; Indels 214; Gaps 39;

QY 25 EVIQAQVQENRLNLESESSQGLGYFSDLNQAPVWYTSSTGDISPSSSELENIP 83
 DB 287 EGVMELEDDDDIVESDAKDESGAGETENS-VDFSKYQPTDWTIKIPVIEKYSDERK 345
 QY 84 SENQVFOAINGS-GFIYVKKSPD-----YTFKTSADN--HYTMWVDDQEVINKASN 131
 DB 346 VHQRTSEGCADFSGSVNISVDDSEDESDQAESYSAANEVYHNEHEHLDKELLEDIES 405
 QY 132 SNKIRLEKRGRLYQIKIQYQRENPTKGLDFKL-----YWTDSQNK 171
 DB 171

```
DB 406 SDS-----ESQSAQSESGSEDDFEYKMKNEKSTSEETENTSESROGFAKDAYTK 456
OY 172 KEVISSDNILQPEL-----KOKSSNSRKRKSTACPTVDPDRDNDGIPDSL 216
DB 457 NKVEQGEENDEEPEKDIIRSLQDNFNHNNKSEYSENVLENTDAIYERENG-IND-- 513
OY 217 EYEGYIVDYKNNKTEL----SPWISNHE--KKGLTYKSS-----PEKSTADPY--S 263
DB 514 -VEGYDVYTKSVESDHEHSP--DNLYDLAARALOFQOOSRNSNCPQKEOYSESYLGHS 570
OY 264 DEEKTATGR-IDKNVSPHARPLVAAPYIVHDMENILSKNEDQSONQNDSERTSKNT 322
DB 571 NGSNLSGRBLDES---EQPLPKD-----FTGNNNNLKTDRGDLSSVEI 613
OY 323 STSRHTSEVHGNAEVH-ANTSTRTHTSEVHGNAEVHAVIDHSLSLAGEERTWAEITGL 381
DB 614 EYEVSEKKLDGSTEKEVPLSTDTINNSSLGNEDESIYSLDDAALISENLIDVLEML 673
OY 382 NTA-----DPAFLMANIRYVNTGAPYIVVLPPTSLVIGKQTLAT 422
DB 674 KTTPKYEVVISESVYSTSEYEDNTVAMPPOVEY---TSPFMD-PPNSD---ND---D 721
OY 423 IAKENQLOSLILAPNNYYSKNNLAPIALNAQODESSTPITM-----NYNQFLELEKT 474
DB 722 YEKKHDLKSTLAA-----LAPATKKDAEFVAGVYKTSCLTSGHTIFHTSKET 773
OY 475 KOLRLDTPDOVGNATYFNENGFRVYDGTGSNNSEVLPOIQTETARIIFNGKDLNVERRI 534
DB 774 KQVS-DLDESTEYVTEPENTG---DENKNQSKNFPVANSKDTEDNTD---EKVF 824
OY 535 AAVNSDPLETTKPDMTLKALKIAGFNEPNGNLOYOGKDTEDPFR-DOQTSON--- 590
DB 825 SAINYTN---VTGSSCCDIETIASNVEE---NLRYCEKDMNEAMSSGDCEVKNQNDG 877
OY 591 IKNOLA-----ELNATNIYVLDKIKLNKANNILIRDKRFHYDNNNIANGADESVKRE 643
DB 878 SKTQISFSDSDPNQESNDNTEFFSSTK-----YKRNSDLEDESKLRE 922
OY 644 AHR-EVIN-----SSTEGLLNIDKIRKILSGYIVETEDTEGLK 682
DB 923 LKRAEVVLDKLEDESESDYEQYADPEPNDGSENNENIYKGFVK-----DTLGLV 972
OY 683 EYINDRYMDLNTISSLQDQKTFIDFKYDKLPLIYSNRYKYNV 727
DB 973 EPENEKVN-----KVHEETLFEANVSSSVNV 999

RESULT 5
Y338_MYCGE
ID Y338_MYCGE STANDARD: PRT: 1271 AA.
AC P47580;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL LIPOPROTEIN MG338 PRECURSOR.
GN MG338.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN
RC SEQUENCE FROM N.A.
RP STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Usterback T.R., Saudke D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bolt K.F., Hu P.-C., Lincer T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
[2]
```

```
RP SEQUENCE OF 1023-1114 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bolt K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
ANCHOR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE MG307 / MG309 / MG338 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
```

DB 829 IRSMONLT---SKER-----FGYKDLGSVNSTSTMLOHQAOSHTSSVNOULL 877
QY 470 -----ELKTKQLRLDQVYGNL---AFYFNENGRVRVDGSMNSEVLPQOETT 517
DB 878 DLAKKAFFKELDDPNO---DAETKTRFLOALMLLVKNG-----AONKNLLOQAIPIG 928
QY 518 ARIIFN---GKDLNIVERIAAVNPSDPLETTKPDWTLKLEALKIARG-----FNEPNCN 568
DB 929 TRAFVSWYGVYKDN-----PSATVSQKTKSSSTSSANENFNFELON 969
QY 569 LOY-OGKDIETEPDNFDQOTSONIKQALAEALNATNIYVLDK-----IKLAKMNLILND 622
DB 970 PYVTGSEIEMFN---DKOTPIQPDLSL---LESENTYRFTDEFPNNSVALSNKQ-Q-SSSD 1022
QY 623 KRPFDRNNIAVGADESIVYKFAHREYINSGTEGLLINDIKDKIRKL-----SGYIYEID 677
DB 1023 KYTFEGFNGLTINSNOS-----ISTASAGLTOQLPNNNGQLITPAIDKAGALSQYKD 1073
QY 678 TEGLEKVID-----RYDMLNLSLRQDKTFIDF---KKY 710
DB 1074 KFTLMSLIKKTSSDAELNAGELHRSVAVDNNLSRFSRGEPLISFDNKKKF 1127
RESULT 6
SMC3_YEAST
ID SMC3_YEAST STANDARD: PRT: 1230 AA.
AC P47037:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CHROMOSOME SEGREGATION PROTEIN SMC3 (DA-BOX PROTEIN SMC3).
GN SMC3 OR YJ1074C OR J1049.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H303;
RX MEDLINE=97474309; PubMed=9335333;
RA Michaelis C., Ciosk R., Nasmyth K.;
RT "Cohesins: chromosomal proteins that prevent premature separation of
RT sister chromatids."
RL Cell 91:35-45(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Rose M., Koetter P., Entian K.D.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Sor F.T.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y14278; CAA74655.1; -
CC DR EMBL: Z49349; CAA89366.1; -
CC DR EMBL: X88851; CAA61313.1; -
CC DR SGD: S0003610; SMC3.
CC DR InterPro: IPR003439; ABC_transport.

DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02463; SMC_N; 1.
DR Pfam: PF02463; SMC_N; 1.
KW Mitosis; ATP-binding; Coiled coil; Nuclear protein.
FT NP_BIND 32 39
FT DOMAIN 172 482
FT DOMAIN 685 1041
FT COILED COIL (POTENTIAL).
FT COILED COIL (POTENTIAL).
SQ SEQUENCE 1230 AA; 141336 MW; B1520887780341F CRC64;
Query Match 4.5%; Score 171; DB 1; Length 1230;
Best Local Similarity 20.2%; Pred. No. 0.25;
Matches 174; Conservative 131; Mismatches 335; Indels 220; Gaps 36;
QY 22 GNLEVIQAEVKQKPNRLNLSSESSOGILGYFEDLNFQAPMVY---SSTTGDSITSSSEL 79
DB 201 GNLSKLSSEMEQERKELEKYNELERNKRIYQFLYRELVEINQMERLDGDYN----- 254
QY 80 ENIPSENOYEQ-----SATWSGFIVKKSDEYTFATSAADNHYTMVWDQ 123
DB 255 NIYSSSEYIQLDKREDMDQYSKLLSI-EASLKIKNATDLOQAKRSEISOKLTNY 313
QY 124 EVINK-----ASNSNKRILEKGRLYQIK-IQYQRENPTKGLDFKLYWTDSONKKEVIS 176
DB 314 NVKIKDVQOOIESNEQORNLDSATLKEIKSIIQOROKKSLILPRYOELT---KEEAMV 369
QY 177 SDNLQLEPELKOKSSNBRKRSSTAGPTVPDRDNDGIPDSLEVGYVVDYKAKRTFSPWI 226
DB 370 --KLQASLQOKORDILLLKGEYARFKSKDERDTWIHSEIE-----ELKSS----- 414
QY 237 SNHFE-KKGLTKKSSPEKSTASDPYSDPEKTYGRDKNVSPEARHPLVAAPIYHVM 295
DB 415 QNLELESQLOMDRTSLRKQYSAID---EELEELDSINGPDYTGOL-----EDFDS 463
QY 296 ENILSKNEDQSTONTDSEFRTISKNTSRTTSEYHGAENVANTSTRTTSEYHGN 355
DB 464 ELHLKOKLSESLDTRKELMRKQKQIQTVELTLDVQNO-QNNVNETMRS-----LAN 517
QY 356 AEVHAVAIDHSLSLAGE-----RTMAETNG-----LNTADTARLNA 351
DB 518 GIINVEIETEKLIKSPESVFGTLGELIKVNDKKTGCEVIGGNSLPHIIVDTEETATLM 577
QY 392 NIRYVMTGAPRIYVNLPTSTSLVGLKNQTLATITAKENQLSQILAPNNYPSKMLAPIALN 451
DB 578 NELYRKAGGRVTF--IPLNLSLSDVKKPSSNTTQIQFTPLIKKIYEPREFKA----- 630
QY 452 AODDFSTPTIMYNNQFLELEKTKQLR---LDTDQV-YGNIAF--YNFENGRVRVDGSM 505
DB 631 VKHVFGKTIYVKDLGGGLAKKHKLAATLDDGDRADKRGVLGGYLDQKRTLRLESKLN 690
QY 506 WSEVLPOIQTARIIFPNGDLNVERIAAVNPS-----DPLETT 546
DB 691 LNSRSQHKRLIELDFVRNELNDIDITKIDQVGNIKIKVSNDRSVLTNIEVYRTSLNTK 750
QY 547 KPD-MTLKEALK-IAFENPNNGNLQYQGDITPEPDN---PDQQTSONIKQOLA----- 596
DB 751 KNEKLILLESNLAIILKLELNTNRPFAQEKLNTFENDLLQEPDSSELSKEKELSLTK 810
QY 597 -----ELNATN-----ITYVLDKIKLAKMNLILNDKRFHNRNNAVAG----- 635
DB 811 EISAANKKLITSDALEGITTTID--SLNDELKSLIPQENDELKSKSEVGDATFIFLOD 868
QY 636 -----ADESVYKEAHREYINSGT-----EGILL--N 659
DB 869 ELKELQEKRSVVEKHQBNVAVELGTVOREIESLAEETNNKKLLEKANNQOORLLKRLDN 928
QY 660 IDKDIRKILSGYIYEIDTGKLE-----VINRBYDM-----LN--I 694
DB 929 FOKSVEKTMKTKTTLVVRRELQORIREIGLPEADALVNDPSDITSDOLLQRLNDMNTET 988
QY 695 SSLRQDCK-TFIDPKKYNDK 713

Db 989 SGLNKNRKAENEKKNER 1008

RESULT 7
MYSL_YEAST
ID MYSL_YEAST STANDARD: PRT: 1928 AA.

AC P08964;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN-1 ISOFORM (TYPE II MYOSIN).
GN MYO1 OR YHR023M.

OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
[1]
RN RP
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RX MEDLINE=91088308; PubMed=2263482;
RA Sweeney F.P., Watts F.Z., Pocklington M.J., Orr E.;
RT "The MYO1 gene from Saccharomyces cerevisiae: its complete nucleotide
sequence.";
RL Nucleic Acids Res. 18:7147-7147(1990).
[2]
RN RP
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnson M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Larellle P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
RA Nman M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
VIII.";
RL Science 265:2077-2082(1994).
[3]
RN RP
RP SEQUENCE OF 1-760 FROM N.A.
RC STRAIN-S288C;
RX MEDLINE=88115139; PubMed=3322809;
RA Watts F.Z., Shields G., Orr E.;
RT "The yeast MYO1 gene encoding a myosin-like protein required for cell
division.";
RL EMBO J. 6:3499-3505(1987).
CC -1- FUNCTION: REQUIRED FOR CELL DIVISION.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
DR EMBL: X53947; CAA37894.1; -
DR EMBL: X06187; CAA29550.1; -
DR EMBL: U10399; AAB68872.1; -
DR PIR: S05806; S05806.
DR PIR: S12323; S12323.
DR PIR: S46773; S46773.
DR HSSP: P08799; 1MND.
DR SGD: S0001065; MYO1.
DR InterPro: IPR000048; IQ.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MISC; 1.

DR PROSITE; PS50096; IQ; 1.
KW Myosin; Actin-binding; Coiled coil; Alkylation.
FT DOMAIN 1 793
FT 794 823
FT 856 866
FT NP_BIND 180 187
FT DOMAIN 460 529
FT MOD_RES 703 703
FT 36 36
FT 46 46
FT 59 59
FT 86 86
FT 330 330
FT 343 343
FT 421 426
FT 515 515
FT 529 535
FT 541 541
FT 550 551
FT 573 573
FT 582 582
FT 588 599
FT 599 599
FT 627 632
FT 695 695
FT 736 742
FT 756 756
FT 773 784
FT 793 794
FT 896 896
FT 900 900
FT 906 906
FT 911 911
FT 915 930
FT 934 939
FT 951 953
FT 955 958
FT 1002 1002
FT 1049 1049
FT 1056 1056
FT 1060 1060
FT 1085 1085
FT 1123 1123
FT 1133 1133
FT 1144 1146
FT 1159 1166
FT 1179 1181
FT 1184 1185
FT 1188 1204
FT 1224 1224
FT 1228 1228
FT 1253 1253
FT 1311 1323
FT 1400 1400
FT 1454 1554
FT 1568 1568
FT 1630 1646
FT 1698 1704
FT 1725 1737
FT 1754 1757
FT 1777 1777
FT 1788 1788
FT 1825 1825
FT 1882 1882
FT 1902 1904
COILED COIL (POTENTIAL).
ATP (BY SIMILARITY).
ACTIN-BINDING (BY SIMILARITY).
ALKYLATION (BY SIMILARITY).
K -> I (IN REF. 1).
I -> T (IN REF. 1 AND 3).
V -> S (IN REF. 1 AND 3).
L -> F (IN REF. 1).
MISSING (IN REF. 1 AND 3).
N -> S (IN REF. 1 AND 3).
OAKAFI -> TKLSSL (IN REF. 1).
D -> S (IN REF. 1).
SKGPTG -> ARGHDR (IN REF. 1 AND 3).
D -> V (IN REF. 1 AND 3).
TD -> LM (IN REF. 1).
R -> A (IN REF. 1).
H -> D (IN REF. 1).
EYVEGLSKRK -> NTKKAGYPT (IN REF. 1).
MISSING (IN REF. 3).
EKSSA -> GKNLVC (IN REF. 1 AND 3).
R -> S (IN REF. 1 AND 3).
ENSTTT -> RKNHHD (IN REF. 3).
E -> R (IN REF. 1 AND 3).
NTKLPKAGVLA -> ILTFQKLEYS (IN REF. 1).
KL -> NV (IN REF. 1).
N -> T (IN REF. 1).
N -> NSQIKRININITEPOSTYIGRPVIGCN (IN REF. 1).
N -> I (IN REF. 1).
N -> K (IN REF. 1).
NESLNKYSSESLQ -> RIAKIKPAINIT (IN REF. 1).
DDLWSE -> MTEFL (IN REF. 1).
AON -> RKI (IN REF. 1).
EEAH -> KKLD (IN REF. 1).
S -> C (IN REF. 1).
L -> D (IN REF. 1).
C -> S (IN REF. 1).
M -> I (IN REF. 1).
A -> E (IN REF. 1).
V -> C (IN REF. 1).
L -> S (IN REF. 1).
KSN -> NLI (IN REF. 1).
RETKEDQK -> TRKEQDKE (IN REF. 1).
SKI -> ELKV (IN REF. 1).
LE -> WK (IN REF. 1).
LSQETSLNQVLRKRS -> CHRYKLSILKQKIR (IN REF. 1).
P -> S (IN REF. 1).
E -> Q (IN REF. 1).
E -> Q (IN REF. 1).
PKKESDINKMLE -> LTKSLILTNQNAS (IN REF. 1).
D -> H (IN REF. 1).
SEOLDRLKLESTEROKELLSITIKQKQOFENCMDDLG
NELRRLRHIALKQAEDEVNMAIIRKLTQKQKRLIV
EREKERNDSQMOLEYLE -> P (IN REF. 1).
D -> V (IN REF. 1).
DLKQDHYKRVEMLN -> SEARSLYKSGNVD (IN REF. 1).
MISSING (IN REF. 1).
TLQLOEQNSRNG -> NTTANGTKFKEM (IN REF. 1).
FDDE -> LMM (IN REF. 1).
D -> E (IN REF. 1).
R -> T (IN REF. 1).
S -> D (IN REF. 1).
S -> W (IN REF. 1).
FWK -> NSGRKIDADDL (IN REF. 1).

SEQUENCE 1928 AA; 223634 MW; 6F54C7611F43DC9F CRC64;

Query Match

Best Local Similarity 4.4%; Score 166.5; DB 1; Length 1928;
Matches 162; Conservative 145; Mismatches 291; Indels 247; Gaps 39;

20 STGNLEVIQAEVQENRNLNSES-----SSQGLGYYPSDLPFOAPMVVSTTGDLS 73
DB 894 TVAELENTODLLQEKENLNKNSLNRKVTSSSETLOKOPDDL-----VSEKDEIS 944
QY 74 IPSSSE-ENIPSENOYFQSAIWSGFLKVKKSDEYTFAT--SADNHVTMWDQDEV-INKA 129
DB 945 REKLEVAONLEAHQKIQ-----GLOETIRERATLEKLKSKNNELIKQSLDNCLODSKE 999
QY 130 SNS-----NKILE-----KGRLOYIYQYORENPTEKGLDFKLYWD----- 167
DB 1000 QSSQSLIKESKLEIKRLKDVINSKEEIKSFNDKLSSESSEDDLIKLVLEKNCNIA 1059
QY 168 -----SON--KKEVISDN-----LOLPLOKSSNSRKRSTAGPT 203
DB 1060 MSRLQSLVYENSPLDRSKNEFKKAKALNNQKNSSELKMEKIDNHKKELATES--- 1116
QY 204 VPDRODNGIIPDSLEVEG-YTVDVKNKRTPLSPWISNIHEKGLTKYKSPEKSTASDPY 262
DB 1117 -KORD-----DAVSEHGKITAELKETRI-----QLETKSYNOK--IKEEY 1154
QY 263 SDEKTYGRIDKAVSPEARPLVAAPIYVHVDENITLSNEDOSTQNTDSERTISKNT 322
DB 1155 SNQORET-----KEQEQKRNKSLV-----ESLNDKIKLELAR-----LSQEI 1192
QY 323 STSRTTSEVHGAEVYHANSSTRTHTS-----EYHGNAEVH 359
DB 1193 SLNQLYKRIKSGNS-VETNISTSTRSTYSDDPKEDIIKKYYDLOLATELTRNLENE 1251
QY 360 AVAIDHSLAGERTAAET-MGLNTADTARLNANIRVNTGTAPIYVNLPTTSVLKQKQ 418
DB 1252 ---IEKKNNLSRLRFTETRLASSFEDQKIKQMKKLKLIQDMPSIPDLSTL---NE 1305
QY 419 TLATITAKEKQSLIAPANNIYPSKNIAPIALNAQDDFSPTITMANYOLELEKTKQLR 478
DB 1306 PLDNCPRKESDINKMLLEVDTL--KROLDETTRAHYD-AENASISALHSKFRKIO--GESS 1360
QY 479 LQTDQVYGNATYVNFENGVRVYDTSNMSEVLQIOET--TARIIPNGDMLNVERRIA 535
DB 1361 LSSSDYI-----KLTFEASERKVSLEDKLTMPLRDRTNLPVGDIIKRDSTISKEBEIR 1416
QY 536 AVNPSDPLETTKPDMLTLEKALIAFGFNEPNGNLOY----- 571
DB 1417 YRK-----LENYK-----NESNGKLSQTLTLDRQSKSEALISQDLRLQK 1462
QY 572 -----QGDITFEFNFDOQTSQNTKQNL-----AELNATNIYVLQK 609
DB 1463 DLESTEROKELLSSTIKQOQFENCMDDQGNELRLREHIIHALKQEEVDKMMASITIEK 1522
QY 610 IKL-NAKMLNLINDKRFHVDNRNIAVGADESIVYKAEHREYNSSTEBLLNIDKDIKIL 668
DB 1523 LKQONQKREKLWEREM--ERNDSMOLOETLLE-----LKVQDVAKIL 1565
QY 669 SGYIVE-----IEDTEGLEKVEYINDRYDMLNISLRQDKFTIDFKKYNDKLPYLSNP 721
DB 1566 SDDLHLKRLERLAVEDRSQYTDIENRLKELN-CSLKAETNLNKEKFAETLKYKLETSTINDS 1624
QY 722 NKVYN 726
DB 1625 EAKIS 1629

RESULT 8

USOL_YEAST STANDARD: PRT; 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE INTRACELLULAR PROTEIN TRANSPORT PROTEIN USOL1.
GN USOL1 OR INT1 OR YDLO58W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-X2180-1A;
RX MEDLINE=91185402; Pubmed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.,
RT "A cytoskeleton-related gene, usol1, is required for intracellular
RT protein transport in Saccharomyces cerevisiae."
RL J. Cell Biol. 113:245-260(1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McCellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE. COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X54378; CA38253.1; -;
DR EMBL: L03188; AAB00143.1; -;
DR EMBL: U03668; AAB6659.1; -;
DR PIR: A38455; A38455.
DR HSSP: P80220; 1DIP.
DR SGD: S0002216; USOL1.
DR Interpro: IPR002017; Spectrin.
KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724
FT DOMAIN 725 1790
FT DOMAIN 465 487
FT DOMAIN 991 1790
FT DOMAIN 1172 1786
FT CONFLICT 847 847
FT CONFLICT 924 924
FT CONFLICT 1253 1253
FT CONFLICT 1319 1319
FT CONFLICT 1461 1461
FT CONFLICT 1581 1581
FT CONFLICT 1600 1600
FT CONFLICT 1661 1661
FT CONFLICT 1772 1772
FT CONFLICT 1772 1772
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B21659FD4818 CRC64;
GLOBULAR HEAD.
COILED COIL (POENTIAL).
CHARGED (HYPER-HYDROPHILIC).
DISPENSABLE FOR THE PROTEIN FUNCTION.
ASP/GLU-RICH (ACIDIC).
G->E (IN REF. 2).
E->K (IN REF. 2).
V->I (IN REF. 2).
I->V (IN REF. 2).
N->S (IN REF. 2).
G->S (IN REF. 2).
I->V (IN REF. 2).
R->S (IN REF. 2).
D->DEEDDEE (IN REF. 2).
D->DEEDDEE (IN REF. 2).

Query Match

Best Local Similarity 4.2%; Score 160; DB 1; Length 1790;
Matches 142; Conservative 145; Mismatches 288; Indels 206; Gaps 32;

```
QY 23 NLEVIQAEVQENRLNSESOGGLGYFSDLNFOAPMAYVSTSTGD-----LSI 74
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1023 NIEOLKRTISDLEQKEEIKSDSSKDEYSOILKEKLETATDENVNKLISELTK 1082
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 75 PSSELE-----NIPSNQYFQSAIWGFIKVKKSEY-----TPAASAHNYTM 118
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1083 TREELAEALAVKKNLKLKLESEKALKVEKNEHEKLKEKIQLEKATEYEQOLNS 1142
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 119 WYDDEIVINKANSNKRIRLEKGRLYQIKI-----OY-----ORENTEKGL 159
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1143 LRANLESEKHEHDLAQLK-----YEOLANKEROYNEEISOLNDELSTQOENSIKK 1199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 160 DFKLYW-----TDSQKKEVYSSDNQOLPELKQSSNSKRSKSTAGPYPPDDN 209
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1200 NDELEGEVKAANKSTSEESNKLKSEIDALNQLKELKKN-----ETNE 1243
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 210 DGIPDSLE-VGGYTVDYK---NKRTEFLSPWISNIHEKKGLRYKSSPKWSTASPYSD 265
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1244 ASLEESISVSEYVYKIELDECFNKEVESELD-----KLKASEDKNSKYLELOKES 1298
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 266 EKVYGRIDKNVSPEARHPLVAAPYIVHVMENII-LSKNEDOSTONTSEFTTKISKTST 324
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1299 EKIKELDAKTE-----LKIQLEKITNLSKAKES-----ESELRLKKTISE 1342
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 325 SRTHTSEVGHAAVEAHANTSTSRHTSEVGHAAVEAHVAIDHSLAGERTWAETMGLNTA 384
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1343 ERK-----NAEQLEKLNKNIQIKNOAFEKERKLNNEG 1375
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 385 DPARLANIRVYVNTGAPYVNLPTTSLVGNKOTLATITAKE-----NOLSOILAPNNY 439
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1376 SSTIQEISEKINT-----LEDELIRLOENEN-ELKAKEIDNTRSELEKYSLSNDE 1424
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 440 YPSKNLAPIALNADDFSS--TPITMANYNOPLLEKTKQLRIQDTQVYGNATYVFNENGR 497
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1425 LLEEKQNTIK-SLODEILSYKDKITRNDKELKILSTIRDNKRDLSESLK-----EDLR 1473
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 498 VRVDGSMSEVLPQIOTETAR---IIFNGKDL-----NLVERRIAAVNPSPDLET-TRP 548
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1474 AAQSEKAVVEEGKLKLEBESSKEKALEKSKEMKLESTIESNETELKSS--METIRKS 1531
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 549 DMTLKEALKIKAFGEFNEPNGNLOYOGKDIREFDNFOOTSONIKNQL-----NELNAT 601
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1532 DEKLEQSKSA---PEDIKNLOHEKSDLS-RINSEKEDIEKSKLIEAKSSELETV 1587
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 602 --NIYVLDKIRLANKMILIRDKRFHYDRNNIAVGADESIVKAEHREVINSSTEGLLLN 659
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1588 KOELNNAQEKIRINAEEMTVLKS-----LEDIERELKDAQE---IK 1627
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 660 IDKDIRKILSGIYVIED-----TEGLKEVINDRKDMISSLRQDGKTFIDFKRYN 711
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1628 SNOEKELITSLKLELEEDLDSTQOKAKOSEERRAEVRRKFOVERKSOLEDKAMLETRYN 1687
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 712 D 712
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1688 D 1688
```

```
RT "Conserved protein domains in a myosin heavy chain gene from
RL Dictyostelium discoideum.";
RT Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RC STRAIN-AX2;
RX MEDLINE=90353583; PubMed=2387408;
RA Lucke-Vielmeier D., Schleicher M., Grabatin B., Wippler J.,
RT Gerisch G.;
RT "Replacement of threonine residues by serine and alanine in a
RL phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
RL FEBS Lett. 269:239-243(1990).
RN [3]
RP PHOSPHORYLATION SITES.
RX MEDLINE=88112226; PubMed=2828113;
RA Wagle G., Noegel A., Scheel J., Gerisch G.;
RT "Phosphorylation of threonine residues on cloned fragments of the
RL Dictyostelium myosin heavy chain.";
RL FEBS Lett. 227:71-75(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX MEDLINE=95345066; PubMed=7619795;
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutcliffe K., Holden H.M.,
RT Rayment I.;
RT "X-ray structures of the myosin motor domain of Dictyostelium
RL Biochemistry 34:8960-8972(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX MEDLINE=95345067; PubMed=7619796;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
RL truncated head of Dictyostelium discoidium myosin to 2.7-A
RT resolution.";
RL Biochemistry 34:8973-8981(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=96206189; PubMed=8611530;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II).ADP.vanadate complex of the
RL Dictyostelium discoidium myosin motor domain to 1.9-A resolution.";
RL Biochemistry 35:5404-5417(1996).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX MEDLINE=97452580; PubMed=9305951;
RA Guzik A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT "X-ray structures of the MgADP, MgATPgammaS, and MgADPPNP complexes
RL of the Dictyostelium discoidium myosin motor domain.";
RL Biochemistry 36:11619-11628(1997).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=98070605; PubMed=9405148;
RA Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
RT "X-ray crystal structure and solution fluorescence characterization
RL of Mg.2 (3')-O-(N-methylanthraniloyl) nucleotides bound to the
RT Dictyostelium discoidium myosin motor domain.";
RL J. Mol. Biol. 274:394-407(1997).
RN [9]
RP FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC [1]
CC SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
CC INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
CC LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
CC (MLC-2).
CC [2]
CC SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
CC CORTEX.
CC [3]
CC DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HWM). IT CAN BE FURTHER
CC SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC [4]
CC DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC [5]
CC PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
```


CC THE ACTIN-ACTIVATED ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE
 CC ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
 CC POSITION (688).
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IO DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M14628; AAA33227.1; -.
 DR PIR: A26655; A26655.
 DR PIR: S00250; S00250.
 DR PDB: 1MMA: 03-DEC-97.
 DR PDB: 1MMD: 17-AUG-96.
 DR PDB: 1MNG: 03-DEC-97.
 DR PDB: 1MNN: 03-DEC-97.
 DR PDB: 1MND: 17-AUG-96.
 DR PDB: 1MNE: 17-AUG-96.
 DR PDB: 1YOM: 23-DEC-96.
 DR PDB: 1LYK: 28-JAN-98.
 DR DictyDb: DD01008; mhca.
 DR InterPro: IPR000048; IO.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00612; IO: 1.
 DR Pfam: PF00063; myosin_head: 1.
 DR PRINTS: PRO0193; MYOSIN_HEAVY.
 DR ProDom: PD000355; myosin_head: 1.
 DR SMART: SM00015; IQ: 1.
 DR SMART: SM00242; MYSC: 1.
 DR PROSITE: PS50096; IQ: 1.
 DR Myosin: Coiled coil; Actin-binding; ATP-binding; 3D-structure;
 DR Calmodulin-binding; Methylation; Alkylation; Phosphorylation;
 DR Calmodulin-binding; Methylation; Alkylation; Phosphorylation;
 DR MYOSIN HEAD-LIKE.
 FT DOMAIN 1 761
 FT DOMAIN 762 791
 FT DOMAIN 817 2116
 FT NE_BIND 179 186
 FT DOMAIN 638 660
 FT DOMAIN 738 752
 FT MOD_RES 130 130
 FT MOD_RES 678 678
 FT MOD_RES 1823 1823
 FT MOD_RES 1833 1833
 FT MOD_RES 2029 2029
 SQ SEQUENCE 2116 AA: 243871 MW: 25C3770BB1EE56A1 CRC64;

Query Match 4.2%; Score 159; DB 1; Length 2116;
 Best Local Similarity 19.2%; Pred. No. 2;
 Matches 154; Conservative 125; Mismatches 291; Indels 232; Gaps 38;

QY 2 KRRKVLPLMALSTIIVSSGNELEVIAEYKQENRLNLESESSOGLLGYFSDLNFQAP 61
 DB 1200 OKKKVELDLEKSKQALEEFLAAQALDKLKKLEQLESEVQTQ-----LSBANK- 1249

QY 62 MVTTSSTGOLSPSSLEINPSENQFQSAIMSGFIKKVKSDEYTPATSDNHTVMVD 121
 DB 1250 NVNSDSTNKLH---ETSFNNIKLELEAEOKA-----KQALEKKRLGLESELKLVNDELE 1300

QY 122 DQEVINFAKSNKRIEKGRLYQIKIQORENPTKEGIDFLYMTDSQNKKE----- 173
 DB 1301 EEKKOKSKNEKRYDLER-EYSELKQDLEEVASKKAV-----TEAKNKSESELDEIKR 1353

QY 174 ----VISSDLQIPELKOKSSNSRRKSTSAPTVPDRDNDGIPDSLEVEGYTVDKNR 229
 DB 1354 QYADVSSRDSVQOLTKLQAKNDELNNTA-----EAGGQLDRAERS-----KKKA 1400

QY 230 TF-LSPMISNHEKKGLTKYKSPKKSTASDPISDEKVTGRID--KNVSPARHPLVA 286

DB 1401 EFDLEAVAKNLEEE---TAKVYKAERAKKAAE---TDVYSTSELDADKANSSQ----- 1449
 QY 287 AYPVHVDMENIILSKNEDS-----TQNTDETRTISKNTSTRTSHTEVHGAENVAN 341
 DB 1450 ----YVQIKRL-----NEELSELKSVLEADERCSAIAKAKTASLESJLKEIDAANN 1500

QY 342 TSTSRTHTEVHGAENVANVAIDHSLSLAGERTYAEITMGTLTADTARLANAIRVNTGTA 401
 DB 1501 AKAAERKSK---ELEVAVALLESLE-----DKSG-----TVNVEPIRKDA 1540

QY 402 PIYVNLPTTSLVLCNKQTLATIKAKENQLSOLAPNNYPSKNLAPIALNADPFSSPI 461
 DB 1541 EIDDLRARLD-----REFESRIKDEDK-----KN----- 1565

QY 462 TMYNQFLELEKTKQLRLTDQYGNATVYFENGVRVDTGSMVSELPQIQ-ETTARI 520
 DB 1566 --TRKQFADLE-----AKVEAQREVVYID----RLKKKLESIDIIDLTOLDTETKSR 1613

QY 521 -IFNGK--DNLVYERITAAVNPSPLETTKPDMLTKALKAIACGFNPNGNLQYOGKDI 576
 DB 1614 KIEKSKKKLEQTLAERRAAEGSSKAAD-----EEIRQVW-----OEY 1652

QY 577 TEFDNFNQDQ-----TSQNIKNQLAELATNTITYVLDPKIKLNKAMNIIIDKRFHYDRN 630
 DB 1653 DELRAQLDSERAALNASKKIKSLVAF-----VDEVAEQLEDEFLAKDLKLVKAKRA 1703

QY 631 --NIAGADESVYKAEHREYINSTEGLILINDIRKILSGYVEIE-----DEKKYDK 676
 DB 1704 LEVELLEVRQDLEEEEDRSLEDSKRLTTEVE-DIKK---KYDAVEQYNTKDEAKKK 1759

QY 677 ---DTEGLKEVINDRYDMLNITS-----LRQDKTFT-----DEKKYDK 713
 DB 1760 LTDVDVTLKQLEDEKKNLSESEAKKRLSEENEDFLAKLDAEYKNSRAEKDRKYEKD 1819

QY 714 LPIYISNPYKVNYAVTKENP 735
 DB 1820 L----KDTKYKLNDDEAATKTQT 1837

RESULT 10
 MDS3_YEAST STANDARD; PRT; 1487 AA.
 ID MDS3_YEAST
 AC P53094;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE MDS3 PROTEIN (MCK1 DOSAGE SUPPRESSOR 3).
 GN MDS3 OR YGL197W OR G1307.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_Taxid=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1679;
 RX MEDLINE=97197971; PubMed=9046087;
 RA Coglielina M., Klima R., Bertani I., Delneri D., Zaccaria P.,
 RA Bruschi C.V.;
 RT "Sequencing of a 40.5 kb fragment located on the left arm of
 RT chromosome VII from Saccharomyces cerevisiae.";
 RT Yeast 13:55-64(1997).
 RL [2]
 RN CHARACTERIZATION.
 RP Li M.B., Neigeborn L.;
 RA Unpublished observations (xxx-1997).
 CC -1- FUNCTION: NOT KNOWN; NEGATIVE REGULATOR OF EARLY MEIOTIC GENE
 CC EXPRESSION.
 CC -1- SIMILARITY: TO YEAST YER132C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X91837; CA62947.1; -
DR EMBL: 272719; CA96909.1; -
DR SGD: S0003165; MDS3.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF01344; Kelch; 2.
DR Melos1s.
KW SEQUENCE 1487 AA; 167073 MW; 768AFB8B796E447 CRC64;

Query Match 4.2%; Score 158.5; DB 1; Length 1487;
Best Local Similarity 19.1%; Pred. No. 1.3;
Matches 170; Conservative 145; Mismatches 318; Indels 255; Gaps 42;

14 STILVSTGNLEVIAEAKQENRLNESSSOGLGYFSDLPNQAQVNTSS-----67
553 STLIKFSPISSNSSKAAVQEGRL-----SSGSLDNF-EKNF--PIFARTVSEAQN 643
68 TTGDSIPSELENIPEMVFQSAIWSGFIKVKKSEDTFATSDNHYTMV-----120
644 TQPOVANADAKAPNTPS-----TSDEPSSSSGSLDSTPHYQRNDEE 686
121 DDQEVIN--KASNSKIRLEKGLYQ--IKIYOENPTEKGLDKLYWTD--ONKKE 173
687 DDEDEVSEKPYKSNISI-----YRPIKTESSSTSSNGMIFRVPKEKAATVSNTE 739
174 VISSNLIQLELKOKSSN-----SRKRRSTAGPTVDRNDGIP---213
740 ALLESNLSIQEISRRSSIPSEGLRSSISEAEHORASHPLTSPLEFDSGTPCGK 799
214 DSLEVEGYTVDKNKRTEFSP-----WISNIHEKGLTKYKSPKMGWSTASDP 261
800 QLOQLQOHT--IQNPHNLSPRRFSARSASISYVSSSDRAG-----NSISSRT-SDS 851
262 YSDFEKTGRIDKNVSPEARHP-----283
852 FGT-PPVLGVLVPLPQTRREPNEPPPCPAMSTGNTFRSNTLIDYHMSKASPFSSRR 910
284 -----LVAAPIVHDMENIILSKNEDOSTQVTDSETRISKNTSTSRH 328
911 SSHIRRSSTPETENAFSAKTPASLDGQMLGSLKESQOTQPRMNSFPANETIQTP 970
329 TSEVGNAEVHANTSTSRHT--SEVHGN--AEVHAVAIDHSTL-----AGERTMAETMGL 381
971 TS-----SNMEMSROSVTSTNTPDSLOSNFALEPLLPRLPSLYMPTSTVAFAPFEFTT 1027
382 NTADTARLANIRIVNTGTAPIYVNLPTTSLVLRGNQTLATIKARENDLSQILLAPNNYYP 441
1028 GOVNSKWLALPALDLVAKIYEIRLKYKLIL--EVLYSILAKKEESLICTSLMBET 1084
442 SKNLAPILANQDDFSSTPIWM--NYNOFLELEKT-----KLRIDTDOV 484
1085 FRTKILNSYKGDEKNTNYLISNDYQELKLKLVLENIDNGYDPLLKRSQAOSST 1144
485 YGNATYFENGRVAVDTSNMSEVLPQIOETTARIIFNG-----KDLNL 529
1145 QESSSSANGE-----KTAAGASLET-----SSTNVPTVFAGRDRSHNSVSGICFNSKMI 1196
530 VERRIAAVNPSPDLETTYTPDMTLKALKIATFG--FNEPNGMLQYQGDITEFDENFDQ 586
1197 QGSRSTSGFS-----PRVAKMSLSKEIDPKTFYE-----EYEPEKGSFDDNDQ 1244
587 TSQNTKN-QLAELN-----ATNIYT-----VLDKILNAKMLLIDKREH 626
1245 TNIGSFNLHLPDMNGSISSTNSISSDLEKEEQLOLLEIERDSAEILLDAR- 1303
627 YDRNNIAVGADESIVKAEHREYVNSSTEGLLINIKDIRKILSGYIVEDEDEGLKEYIN 686
1304 ---RNM-----EDDKVTKD-----ISNDKRNITLPHKKNMLKAKEG--KETRDVREEEEDP 1350

QY 687 DRYDMLNITSLRQDKTFIDFKYNDKL-PLYISN--PNYKVNYVAYT 731
1351 FGLGMLSLKKIRKEAK-HVD--KVDSVDPLRKSSAFQSPIRATGST 1395

RESULT 11
FAT1_SCHPO
ID FAT1_SCHPO STANDARD; PRT: 1385 AA.
AC 013735; 09JUT10;
DT 15-JUL-1998 (Rel. 36, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ACTIN INTERACTING PROTEIN 3 HOMOLOG.
GN FAT1 OR SPAC15A10.15 OR SPAC15EL.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetales;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE OF 1-1033 FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1023-1385 FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Lyne M.H.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP IDENTIFICATION, AND GENE NAME.
RX MEDLINE=20143585; PubMed=10679021;
RA Jin H., Amberg D.C.;
RT "The secretory pathway mediates localization of the cell polarity
regulator Alp3p/Buddp.".
RL Mol. Biol. Cell 11:647-661(2000).
CC -!- FUNCTION: INVOLVED IN THE ORGANIZATION AND/OR FUNCTION OF THE
ACTIN CYTOSKELETON.
CC -!- SIMILARITY: TO YEAST BUD6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z97208; CAB10112.1; -
DR EMBL: AL109770; CAB52420.1; -
KW Coiled coil; Cytoskeleton.
FT DOMAIN 1009 1096 COILED COIL (POTENTIAL).
SQ SEQUENCE 1385 AA; 154325 MW; ABB3D40CCAF7537 CRC64;

Query Match 4.2%; Score 158; DB 1; Length 1385;
Best Local Similarity 18.1%; Pred. No. 1.3;
Matches 141; Conservative 129; Mismatches 282; Indels 228; Gaps 30;
55 DLNQPAMVYVTSSTGDL-----SIPSELENIPEMVFQSAIWS-----GFIKVKKS 103
326 DRKFSPKLRLRPSTLRSIDPCTPTSLKSPSLKSSPS--SFVQKDVYSRSNSLRISQANRS 383
104 DEYFATSADNHYTMVDDQEVINKAS-----NSN 133
384 NVFPGATD---VTRSVSDHRLSSSTINDGEVAPPLDQRSRTISSPNSPLSATVLPSTT 439
134 KIRLEKGLYQIKIYOSEN-----PTEKGLDFKL-----YMTDSQMKK 172
440 PILLRGSSSTISVAKKQFNMDGSLTNSPNSIRETEYVASFKEJLADAEVETDATGQR 499
173 EVI-----SSDNLQLELKOKSS--NSRKRRTSAGPTVDRNDGIP---PDSL 216

QY 390 NANIRYV-----NTGAPIVNLPPTSIV-----LG 415
DB 371 EALLVYLPCLLEDMSMNTTIIHIIITSGMTGCSAAAKYVILDMGWIKRNPISQ 430
QY 416 KQOTLATIKAKENOSQILAPNNYPSKRLAPIALNAODFSPTPTMYNOLFLEKTK 475
DB 431 TNEKSKPRDKNGEKRSILLEN--LDLKWITANMLNADNSDTCINT----- 476
QY 476 QLRLDIDOVYGNIAIYNEE-----NGRVRVDJGSNMSEVLPOIETTARIIFNGKDLN 528
DB 477 AARLGNISIVDALDYGADPFIANKSGLRPVDEGAGTS---KLQNT-----NGDENS 526
QY 529 -LVE-----RRIAAVNPSPDLETTKPDMTLKAALKIATGFENPNGLQYQK- 574
DB 527 KVVSKGDYDQGNKRAKRIKRSOLKNPPTSLINDVQMLN-SISKDYENETVOYNEKL 585
QY 575 DTERDFNPDQOTSONIKQOLAEI-NATNIYTVLDKIKLNAKMLILIRKRFHYDRNNIA 633
DB 586 EKLHLELNEQREELANSRQOLANVQOLKDEYSLMQDLTNLKAIEEESTREESKKLG 645
QY 634 VGADES-----VKEAHR-EVINSSTEG-LLNIDKDIRKILSGYIIEDETEGL 681
DB 646 IADSSGIDMSSEYDADEPFKEVFLSDLEDKLOKNVEGDISKLL-----EASKEGI 700
QY 682 KEVINRDRMDLNISSLRQDG---KTFIDPKKYNDK 713
DB 701 MEQIRNQLPAERIKQSMLEPPTVLLKARINAYKRNDK 735

RESULT 13

FNBA_STAAU STANDARD: PRT: 1018 AA.

AC P14738;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FIBRONECTIN-BINDING PROTEIN PRECURSOR (FNBP).
OS FNBA.
OC Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTC 8325-4;
RX MEDLINE=89098998; PubMed=2521391;
RA Sinaas C., Raucel G., Joensen K., Lindgren P.-E.,
RA Antharamiah G.M., Hoeck M., Lindberg M.:
RT "Nucleotide sequence of the gene for a fibronectin-binding protein
RT from Staphylococcus aureus: use of this peptide sequence in the
RT synthesis of biologically active peptides."
RL Proc. Natl. Acad. Sci. U.S.A. 86:699-703(1989).
-I- FUNCTION: THE ABILITY OF BACTERIA TO BIND FIBRONECTIN HAS BEEN
PROPOSED AS A VIRULENCE FACTOR ENABLING BACTERIA TO COLONIZE
WOUND TISSUES AND BLOOD CLOTS. BINDING OF PLASMA FIBRONECTIN TO
THE BACTERIAL SURFACE MIGHT BLOCK ADHESION RECEPTORS ON S.AUREUS,
THUS REPRESENTING AN IMPORTANT DEFENSE MECHANISM AGAINST TISSUE
INVASION.
-I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
-I- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04151: AAA26632.1: -
DR InterPro: IPR001899: Gram_pos_anchor.
DR Pfam: PF00746: Gram_pos_anchor: 1.

DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
KW Signal; Repeat; Cell wall; Transmembrane.
FT SIGNAL 1 36
FT CHAIN 37 1018
FT DOMAIN 37 993
FT TRANSMEM 994 1009
FT DOMAIN 1010 1018
FT REPEAT 545 574
FT REPEAT 575 604
FT DOMAIN 745 878
FT REPEAT 745 782
FT REPEAT 783 820
FT REPEAT 821 859
FT REPEAT 860 878
FT DOMAIN 879 948
FT REPEAT 879 892
FT REPEAT 893 906
FT REPEAT 907 920
FT REPEAT 921 934
FT REPEAT 935 948
FT DOMAIN 982 987
SQ SEQUENCE 1018 AA; 111780 MW; 58175E0020B81F1F CRC64;

Query Match 4.1%; Score 155.5; DB 1; Length 1018;
Best Local Similarity 19.4%; Pred. NO. 1.1;
Matches 165; Conservative 116; Mismatches 311; Indels 257; Gaps 38;
QY 23 NLEVIOAE-VKQENRLNESSESSQGLIGYFS-DLNFQAPVAVTSSTGDISI---PS 76
DB 109 NIETVEEYVKEEAKPOVETQOSODNSGDQROVLD---TPKATQONQVAETQVEVQPR 165
QY 77 SLENIPEENQYFOGAINSGFIKKKSDPYTATGAD--NHTMVAVDQEVYINKSNSNK 134
DB 166 TASESKPRVTRNSADVA-----EAKESNAKVEGTDTVTSKYVEIGSTEGHN--NTNK 216
QY 135 IRLKGRLYQIKIYOORENPTEKG--LDPKLYWTSQNKKEVISSDNLQPLKOKSSNS 192
DB 217 VERPHAGQRAVLKXKLFENGLHQGDYFPDTL---SNVNTHTGVSATARKVPEIKKGSVYM 272
QY 193 RKKRSTAGPYVPDRDNDGIPDSLEY-----EGYTVDVKKKRPFLSWMISNIHEK 242
DB 273 ATGEVLEGKIRYFTFND-IEDKVDVTALEINLFIIDPKTVQTNQNGQITSTLNEQRTSK 331
QY 243 KDLTRYK-----SSPEKWTASDPYSD--FEK-----VTGRIDKNVSE 279
DB 332 ELDDVYTKDQIGTYANLNGSITETFKANKRFSHVAFIRPNNGKTSVYVYTGILMGKSNON 391
QY 280 ARHPLVAAYPIVHVDMENTIISKNDQSTONTSETRTISKNTSTSRHTS---EVHGNA 336
DB 392 GNQPVRIFFE-----YLGNND-----IAKSIVANTTDITSFKKEVTSNM 430
QY 337 EVHANTSTSRHTSEVHGNAEVAHAVALDHSLSLAERTMAETMGANTADTARLANI--- 393
DB 431 SCNLMLQNNGSYSLNIEMDKYVVHYD---GEY-----LNGTDEVDFTQMWCH 477
QY 394 -----RYVNGTAPITAVNLPPTSVLGKNQTLATIKAKENOLSLAPNNYF----- 441
DB 478 PEOLKYKYDRG---YLTITWNGILVYSN-----KANGNEKNPITQNNFEYKEEDTI 527
QY 442 -----SKNLAPIALNAODFSPTPTMYNOFLE-----LEKTKQLRLD- 480
DB 528 KETLGGQYDKNL---VTYVEEYDSSTIDIDHTAIDGGGVDGYIETIETDSALDI 584
QY 481 -----TDQYGNIAIY-----NFEEN-----GRVAVDTGS 504
DB 585 DYHTAVDSAGHVGGYTESSESNPIDEEESTHNSKHADVVEYEDTTPNGCGQVTTES 644
QY 505 NMSEVLPQIOETTARIIFNGKDLNLVERRIAIVNPSDPLETTKPRMTLKEALKIATGFENE 564
DB 645 N-----LVEPDESTKGIVTG-----AVSDHTTVEDTKKEYTTESNLIELVDLPE 689

```

Oy 565 PGNLQYOGKRIETEDDFEEDQDOTSGNINQOALNATNYVLDKIKTNAMKNL----- 619
Db 690 EHGOAQOGVEEITKNNHHISH-----SGLTENGHGHNIVIEETIEENSHYDIKSELGY 742
Oy 620 -----TRDKRFHYDRNNINAVGADSEVYKEAH-----REVINSTE----- 654
Db 743 EGGONGSGNQSEEDPEEDKPKYEGCGNITVDIDFDSVPDIGHQNGKNGQSEEDTEKDKPKY 802
Oy 655 ---GLLINDDKIRILSGY-----IYAEIETEGEIKVEYINDRYDMLNLSLRQDSKTFID 706
Db 803 EHGGITIDIDEDSVPHIHGFKHTEIIE-EDTNKDK-----PSYQFGGHSNVD 849
Oy 707 FKRYNDKLP 715
Db 850 FEE-DTLP 856

RESULT 14
ID IGAL_HAEIN STANDARD; PRT; 1541 AA.
AC PA2782;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE IMMUNOGLOBULIN A1 PROTEASE PRECURSOR (EC 3.4.21.72) (IGAL PROTEASE).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK368 / SEROTYPE B.
RX MEDLINE=68379374; PubMed=2506130;
Rt Poulsen K., Brandt J., Hjorth J.B., Theegersen H.C., Kilian M.;
Rt "Cloning and sequencing of the immunoglobulin A1 protease gene (iga)
Rt of Haemophilus influenzae serotype b.";
Rt Infect. Immun. 57:3097-3105(1989).
RL [2]
RN RP
RP MUTAGENESIS OF SER-288.
RC STRAIN=HK368 / SEROTYPE B.
RX MEDLINE=92234949; PubMed=1373717;
Rt Poulsen K., Relhoidt J., Kilian M.;
Rt "A comparative genetic study of serologically distinct Haemophilus
Rt influenzae type 1 immunoglobulin A1 proteases.";
Rt J. Bacteriol. 174:2913-2921(1992).
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: CLEAVAGE OF IMMUNOGLOBULIN A MOLECULES AT
CC CERTAIN PRO-1-XNA BONDS IN THE HINGE REGION. NO SMALL MOLECULE
CC SUBSTRATES ARE KNOWN.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PERTIDASE FAMILY S6 (SERINE PROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X64357; CAA45708.1; -.
DR EMBL: M87492; AAA24969.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR000710; IGA_S6.
DR Pfam: PF02395; IGAL1.1.

```

RP SEQUENCE FROM N.A.
RC STRAIN-HK393 / NCTC 8467 / SEROTYPE B;
RX MEDLINE-92234949; PubMed-1373717;
RA Poulsen K., Reinholdt J., Killian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: CLEAVAGE OF IMMUNOGLOBULIN A MOLECULES AT
CC CERTAIN PRO-1-XNA BONDS IN THE HINGE REGION. NO SMALL MOLECULE
CC SUBSTRATES ARE KNOWN.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M87490; AAA24967.1; .
DR MEROPS: S06.001; .
DR InterPro: IPR000710; IGA_S6.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PR00921; IGASERPTASE.
KM Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1012 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1013 1545 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 292 292 PROBABLE.
SQ SEQUENCE 1545 AA; 170627 MW; 3EDD75398BF6D478 CRC64;

Query Match 4.1%; Score 155.5; DB 1; Length 1545;

Best Local Similarity 20.1%; Pred. No. 2;

Matches 129; Conservative 98; Mismatches 225; Indels 191; Gaps 33;

QY 21 TGNLEVYQAEVQENRLNESESS-----QGLG-----YFSDL-NFOA-PMVY 64
DB 861 TGNSDVHOLDLNGHILNLSADNSNNVTKYNTLVNLSLNGSGFYLLTDLNKGDKVYV 920
QY 65 TSSGTGDLSPSELENTISENOYFOSAIGWGFYKKSDEYTFATSDNHVTM--WVDD 122
DB 921 TKSATGNFTLOYADKTGEPNHNL-----TLFDASKAQRDHLNLSLVGNFTVDIGAM--- 971
QY 123 QEVINKASNSKIRLEKGRLYQIKIOYORENPTEKGLDKLWTSQNKKEVISSDNIOL 182
DB 972 -----KYLIRNNNGRYDLNPEVEKRNOT-----VDTTN--ITTPNNIOA 1009
QY 183 PELKOKSSNSRKRSTSA-----GPTVPDRDNDGIPDSLEVGTYVDVKNKRTFLSPWIS 237
DB 1010 DVPSPVSNNEELARVDEAVPPAPATPSETTETVAENSKQESKTYE-KNEQ----- 1060
QY 238 NIHEKKGLTKYSSPEKMWSTADPYSDFEKVGRIDKNVSPKARHPLVAAYPIVHVDMEN 297
DB 1061 -----DATEYTAQ-NREYAKKAKSNVKA-----NTQTNE 1088
QY 298 IILSKNEDOSTQNTDS-ETRTISKNTSTSRHTSEVHGNAEYHANTSTRTHTSEVHGNA 356
DB 1089 VAQSGSETKETQTTETKETATYEKE-EKAKVEKTEQEVKVTQSOVSPKQESQETVQPOA 1147
QY 357 EYHAVAIDHSLSLAGERTWAEFTMGLTADTAR-----LNANIRY-----VNTGTAPIYN 405
DB 1148 E-PARENDEPTVAIKEPQSGTNT-----TADTQPAKETSSNVQEPYESTTVNTGNSVVEN 1202

QY 406 VLPTTSVLGKQNTLATIKAKENQSLIAPNNYPS-----KNLAPIALNAD----- 454
DB 1203 -----PENTTPATTPQPTVNSESSNPKRNRHRSVSVPHNVEPATTTSSNDRSTVAL 1253
QY 455 -DESSTPITMNTN-----QFLEKTYQLRIDTQV-YGNATAYNFENGRRVDT 502
DB 1254 CDLTST-----NTNAVLSDARAKAOFVALVNGKAVSHISQLEMMNNGQYV----- 1300
QY 503 GSNWSEVLQIOETTARILFNKRDNLVERRIAAVNPSDPLETKRDMTLKALKIARGF 562
DB 1301 ---W-----VSNTSKNNKYSSQY---RRFSSTQTOGLW---DOTISNNVQGVDF 1344
QY 563 N-EPNGNLQYGGKDLTEFDFNFDQOTSQNIKQALAEINATNIIY 604
DB 1345 TYVRNSN-----NFDKATS---KNTLAQVNFYSKY 1371

Search completed: December 2, 2001, 13:52:29
Job time: 384 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:51:18 ; Search time 72.22 Seconds

(Without alignments)
775.246 Million cell updates/sec

Title: US-09-747-521-4_COPY_1_735

Perfect score: 3765

Sequence: 1 MKRRKVLPLMALSTILVSS.....LYISNPKNKVAVTKENT 735

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

PIR68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3652	97.0	764	2	I39934
2	877.5	23.3	875	2	I40862
3	211.5	5.6	192	2	I39933
4	211.5	5.6	204	2	G59104
5	190.5	5.1	4688	2	F82885
6	187	5.0	1072	2	A86827
7	183.5	4.9	2178	2	S55605
8	183	4.9	1465	2	S31262
9	183	4.9	1803	2	S56894
10	179.5	4.8	1051	2	T18351
11	179.5	4.8	1365	2	T30822
12	177.5	4.7	1467	2	PC1253
13	177	4.7	1658	2	S55101
14	176	4.7	1193	2	S52618
15	175	4.6	1802	2	S52611
16	174.5	4.6	4919	2	T31105
17	173.5	4.6	2269	2	T28677
18	173	4.6	1302	1	JC6009
19	172	4.6	1271	2	D64237
20	172	4.6	1939	2	T18372
21	171.5	4.6	2529	2	B64635
22	171	4.5	1230	2	S56850
23	170.5	4.5	1837	2	T41023
24	170	4.5	1315	2	T28679
25	169.5	4.5	1620	2	S61535
26	168	4.5	2523	2	T18477
27	168	4.5	3724	2	T18427
28	166.5	4.4	1115	2	T41342
29	166.5	4.4	1230	2	E64664

30	166.5	4.4	1928	2	S46773	myosin heavy chain
31	166.5	4.4	2401	2	T28676	rhodopy protein -
32	166.5	4.4	2526	2	T20531	hypothetical prote
33	166	4.4	821	2	S67087	hypothetical prote
34	164.5	4.4	1553	2	T18502	hypothetical prote
35	163.5	4.3	1014	2	H71602	protein with DnaI
36	161.5	4.3	1595	2	T31082	endo-1,4-beta-xyla
37	161.5	4.3	4152	2	T31102	filamentous hemag
38	161	4.3	1538	2	T29095	cardiac muscle fac
39	161	4.3	2399	2	H71879	toxin-like outer m
40	160.5	4.3	888	2	A38539	p101 protein precu
41	160.5	4.3	1979	2	C71622	hypothetical prote
42	160.5	4.3	2340	2	B71704	cell surface antig
43	160	4.2	1790	2	S67593	transport protein
44	159	4.2	1516	2	E71619	RAD2 endonuclease
45	159	4.2	2116	2	A26655	myosin heavy chain

ALIGNMENTS

```
RESULT 1
I39934
C:Species: Bacillus anthracis
C>Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 01-Dec-2000
C:Accession: I39934; S69160; F59104
R:Melkos, S.L.; Lowe, J.R.; Eden-McCutchan, F.; Vodkin, M.; Leppla, S.H.; Schmidt, J.
Gene 69, 287-300, 1988
A:Title: Sequence and analysis of the DNA encoding protective antigen of Bacillus ant
A:Reference number: I39933; MUID:89172073
A:Accession: I39934
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-764 <RES>
A:Cross-references: GB:M2589; NID:g143280; PIDN:AAA22637.1; PID:g143282
R:Fieldman, T.C.; Gordon, V.M.; Leppla, S.H.; Klimpel, K.R.; Birch, N.P.; Loh, Y.P.
Arch. Biochem. Biophys. 316, 5-13, 1995
A:Title: In vitro processing of anthrax toxin protective antigen by recombinant PC1 (
A:Reference number: S69160; MUID:95142670
A:Accession: S69160
A:Molecule type: protein
A:Residues: 197-202 <FR1>
R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koeh
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pX01, the large Bacillus anthracis plasmid harb
A:Reference number: A59091; MUID:99445483
A:Accession: F59104
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-313, 'Q', 315-764 <OKI>
A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32414.1; PID:g4894326
A:Experimental source: strain Sterne
A>Note: similar to anthrax toxin moiety, protective antigen, pagA formerly pag, plasm
C:Genetics:
A:Gene: pX01-110
A:Genome: plasmid
C:Function:
A:Description: three component exotoxin; protective antigen binds to receptors on the
Y active components edema factor or lethal factor; the complex is internalized by rec
C:Keywords: exotoxin
F:1-29/Domain: signal sequence #status predicted <IG>
F:30-196/Domain: propeptide #status predicted <PRO>
F:197-202/Product: protective antigen #status experimental <MAT>
```

```
Query Match 97.0%; Score 3652; DB 2; Length 764;
Best Local Similarity 97.6%; Pred. No. 8e-182;
Matches 717; Conservative 2; Mismatches 16; Indels 0; Gaps 0;
OY 1 MKRRKVLPLMALSTILVSSGTGNEVIOAEVKQENRLNSESSESSSGGLGYFSDLNQOA 60
DB 1 MKRRKVLPLMALSTILVSSGTGNEVIOAEVKQENRLNSESSESSSGGLGYFSDLNQOA 60
```

QY	61	PMVWTSSTTGOLSTSPSSLEKMIPEENQYFQSAIWSGFIKXKDEYTPATSAHNHTMAY	120
Dp	61	PMVWTSSTTGOLSTSPSSLEKMIPEENQYFQSAIWSGFIKXKDEYTPATSAHNHTMAY	120
QY	121	DDQEVYKXASNSNKRILEKGRVQIKIQYORENTEKGLDFKLYWDSQNRKEVSSDNL	180
Dp	121	DDQEVYKXASNSNKRILEKGRVQIKIQYORENTEKGLDFKLYWDSQNRKEVSSDNL	180
QY	181	QLPELKQSSNRKRKSTASAGPYDPBDNDNGIPDSLEVEGYTVDVYKNRFLPSWTSNH	240
Dp	181	QLPELKQSSNRKRKSTASAGPYDPBDNDNGIPDSLEVEGYTVDVYKNRFLPSWTSNH	240
QY	241	EKKGLTKYKSSPEKWSASTADPYSDFEKYVTRIDKNVSPEARHPLVAAYPVIVHDMENIL	300
Dp	241	EKKGLTKYKSSPEKWSASTADPYSDFEKYVTRIDKNVSPEARHPLVAAYPVIVHDMENIL	300
QY	301	SKNEDQSTQNTDSTRTISKRTSTSRTHSEVHGNAEYVHANTSTSRTHSEVHGNAEYHA	360
Dp	301	SKNEDQSTQNTDSTRTISKRTSTSRTHSEVHGNAEYHANTSTSRTHSEVHGNAEYHA	360
QY	361	VAIDHSLSLAGERTMAETMGLNTADTARLANANRYVNTGAPRYVNLPTTSLVIGKNQVL	420
Dp	361	VAIDHSLSLAGERTMAETMGLNTADTARLANANRYVNTGAPRYVNLPTTSLVIGKNQVL	420
QY	421	ATIRAKENQUSQILAPNNYYPKSNLABIALNAODDPSPTITMNYQOFLELEKTKOLRD	480
Dp	421	ATIRAKENQUSQILAPNNYYPKSNLABIALNAODDPSPTITMNYQOFLELEKTKOLRD	480
QY	481	TDQYVGNATYVNFENGVRVDTGNSMSEVYPOJQETTARIIFNGKOLNVERIAVAVNS	540
Dp	481	TDQYVGNATYVNFENGVRVDTGNSMSEVYPOJQETTARIIFNGKOLNVERIAVAVNS	540
QY	541	DPLETTPKPDMLKEALKIAEGFNEPNCNLQYQKDIITFEPFNDQOTSONIKNOJLAELNA	600
Dp	541	DPLETTPKPDMLKEALKIAEGFNEPNCNLQYQKDIITFEPFNDQOTSONIKNOJLAELNA	600
QY	601	TNITVTVDIKRIKLNKKNMILIRDRFPHDRNNIAVGADESYYKRAHEVINSSTEGLLNI	660
Dp	601	TNITVTVDIKRIKLNKKNMILIRDRFPHDRNNIAVGADESYYKRAHEVINSSTEGLLNI	660
QY	661	DKDRIKILISGYIIEDETEGLEKVIENDRYMMLISSLRQDGKFIIDFKYKNKRLPLYISN	720
Dp	661	DKDRIKILISGYIIEDETEGLEKVIENDRYMMLISSLRQDGKFIIDFKYKNKRLPLYISN	720
QY	721	PNYKVNYYAATYKENT 735	
Dp	721	PNYKVNYYAATYKENT 735	

```

RESULT      2
I40862
iota toxin component Ib - Clostridium perfringens
C:Species: Clostridium perfringens
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Oct-1999
C:Accession: I40862; S42774
R:Perelle, S.; Gilbert, M.; Boquet, P.; Popoff, M.R.
Infect. Immun. 61, 5147-5156, 1993
A>Title: Characterization of Clostridium perfringens iota-toxin genes and expression
A:Reference number: I40861; MUID:94041637
A:Accession: I40862
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-875 <RES>
A:Cross-references: EMBL:X73562; NID:g929031; PIDN:CAAS1960.1; PID:g414655

Query Match          23.3%, Score 877.5, DB 2, Length 875;
Best Local Similarity 31.9%; Pred. No. 5,3e-38;
Matches 254; Conservative 132; Mismatches 270; Indels 139; Gaps 33;

OY      15 TIVVSTGNLEY-----IQAEVKQENRLNESSSSQGLGYFSFLNQAPWVVSSTT 69
|::||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

```

Db	15	TAMISOTLSYVAVYAQTTOJNDOTNOKEITENTENLSSNGLMGYFADEHEKDELEMAPJKN	74
OY	70	GDLISPESELENTIPSE-NOYFOASIAWSGFLYAKXSODEYFATASADNHYTMWVDOOEVIK	128
Db	75	GDLFEERKEKVKDKLTLEENSSIKSIRMTGRIPSEDEYILLSTR-NDVLMQJINAGDIK	133
OY	129	ASNSNKRILEKGRLYQJIKIYOOREN-----PTEKGDLFLWYTDSONKEVSISSN	179
Db	134	---FLTKNMKKGAQVYNRIEIDOKNLGSIIDNLSVP-----KLYM-ELNGKTYIPEEN	182
OY	180	LOLPELKOKSSNSRKKRSTAGPTV-----DRDNGJIPDSLEVEG	220
Db	183	LFERDYSKIDEND-----PEIPNNNEFDVREFESAAMEDEDLTDNDINIPDAVEKNG	233
OY	221	XYVUVKKRFRFLSPWISNIEHKKGLRYKXSSPEKMSSTADPVSDEPKYGRIDKNVSPA	280
Db	234	YTI---KDSIAVAKMDSFAE-OGYKRYVSSYLESSTADPPTDYOKAAGSIDIKAIEPA	288
OY	281	RHPLVAAYPIVHVDMENTIILSKNEDOSTONTDESTRTISKNTSRTHSEVYHNAEYHA	340
Db	289	RDLPLVAAYPVYGVGMELIISTNEHSS---DQGTIVSRATINSGTDNTANY--GVYSISA	342
OY	341	-----NTSSTRHTSEVYHNAEYHAVAIDHSISLAGERTWAETMGLTADTARLAN	392
Db	343	GYONGFEGNTTSSYSHTTD-----NSTAPVDS--NGE-SMNGLSINKGESAYIAN	391
OY	393	IRVYNGTGTAIYVULPPTSGVLKNOQTATIRAKENQSOIILAPNNYPSKULAIALNA	452
Db	392	VRIYNGTGTAPMYVYTPPTNLV--DGETLATIRAKODQOINNLSMPEYTPKRLSPALNT	450
OY	453	QDDFSSPTPIPMANTNOPELETEKTKOLRLDPDOVYGNATYVFNFGGRVYDVSMSMEVLPO	512
Db	451	MDQGNALPIINYNOLKJIDSGKQIKLETTQVSGNCTGN-SOGOI-ITEGMSMNSIYQ	508
OY	513	IOETTAIIPINGKDLNVERRIAANVPSDPLETTKPDMLKALAKIAGEFENPGNLOY	571
Db	509	IDVSASIIILD-NGSOTFEFFERRAFAKQGNDEKXT-PEITIGALIKKAFSATK-NEELLYF	565
OY	572	OGKQITE--PDFNPDQOTSONIKNOLAELNATNYVLDKILNKMNLINDKRF--HY	627
Db	566	NGPIDESCCELLFDDNTSEITIEOLUKYDOKKIYV---KLEGMNLIIIVSPYFTNF	621
OY	628	DR--NNIAGADESVVEAREVINSSTEGH-----LNDIDDKIRKISGY-----	671
Db	622	DEVXNNPF--ASMSNIDPKNQDGLQSVANKLSGETKILIPMSLKPKYKRIVFEGYSKDSY	679
OY	672	---IYEIEDTEBGLKEYVINDRYMLNIS-----SLRQDGKFFIDFKKYND	712
Db	680	SNSTIVYNIKSEKORTDYLPEKDYTKFSEYFETTGKQSDSIDIEITLSSGVIFLDMLSITE	739
OY	713	--KLPIYISNPKY 725	
Db	740	LNSTPELIKPEPIKV 754	

RESULT 3
139933
Cryptic protein - Bacillus anthracis
C:Species: Bacillus anthracis
C:Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 15-Oct-1999
C:Accession: I39933
B:WMLKOS, S.L.; Lowe, J.R.; Edén-McCutchan, F.; Vodkin, M.; Lepaja, S.H.; Schmidt, J.
Gene 69, 287-300, 1998
A:Title: Sequence and analysis of the DNA encoding protective antigen of Bacillus anthracis
A:Reference number: I39933; MUID:99172073
A:Accession: I39933
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-192 <RMS>
A:Cross-references: GB:M25589; NID:g143280; PIDN:AAA22636.1; PID:g143281

Query Match 5.6%; Score 211.5; DB 2; Length 192;


```

Query Match          5.6%; Score 211.5; DB 2; Length 204;
Best Local Similarity 34.8%; Pred. No. 0.0002;
Matches 46; Conservative 31; Mismatches 34; Indels 21; Gaps 4;

OY      616 NNILIRKRFHYDRNNINAVGADSEVYKEAHREVINSSTEGLLNIDKDIRKILSGIYEI 675
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db       1 NNIILVRDP-YHNDNGNITGVDDSYLKNAKYKOILMWSSDGYSLNDEVDNALSGYMLQI 59

OY      676 EDTE-----GLKEVINDRYDMLNISLRDGKTFIDEKKYNCKLPLTYSNP 722
        :         :         :         :         :         :         :
Db       60 KKPNSHLTNSPYTTTLACKDSGVGEELRYVL-----DTGFLEDFNKFDENWRSLV-DPG 112

OY      723 YKVNYVAATKEN 734
        | |||||:
Db      113 DDVYVVYAATKED 124

RESULT  5
F82885
hypothetical protein U0482 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: F82885
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mild
A:Reference number: A82870
A:Accession: F82885
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-468 <GLA>
A:Cross-references: GB:AF002145; GB:AF222894; NID:g6899476; PIDN:AAP30894.1; GSPDB:GN001
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: U0482

```

A:Genetic code:SGCC3

Query Match
Best Local Similarity 20.7%; Score 190.5; DB 2; Length 4688;
Matches 168; Conservative 132; Mismatches 341; Indels 171; Gaps 39;

OY 18 VSSGNGNEVLT-QAEYKQDENRLNSESSESOGLCYTYSDLNFQAPRWYTSTTDLSIPS 76
 ::|||:: : : : : : : : : : : : : :
DB 3676 INKTGRKOPELVSVAGLKSNOLEP-----LVVVYYLD-NIHONIDETKRIFKDHW-S 3725

OY 77 SELENIPSENQYFOSAIMSFGPIVKKKSDPYEAFASADNHVMWDOOEVIKAASNKRIR 136
 ||| | : : : : : : : : : : : : :
DB 3726 KEIENLPDYVTMISKHGNKSPDTTTANFEFYELIQ-----DDNDVLNNIDATAYKF 3776

OY 137 LEKGRLYOIKIOYORENPTEKEGLDKLYMTDSQNKKVEISSDNLOPELKOKSSN---S 192
 : : : : : : : : : : : : : : : : :
DB 3777 DEHNINKRKYRIILKEN-----NMWLKGO--DNLN-DETKYLEINIELSK 3820

OY 193 RKKNSTSGAPVPDRD-----NCGISPSLEYEGTYDVAKRKPFIS 233
 : : : : : : : : : : : : : : : : :
DB 3821 PLKPTTNLSVSINDKENISLITETGNPVLYKIOTQNTINDTOQTIVTLTGAVNSK-YNG 3879

OY 234 PWINSIEHKKGLTRYKSSPEKMSASTADEYSPFEKVTCRGIDKNVPEARHPPLVAALPVHV 293
 | : : : : : : : : : : : : : : : : :
DB 3880 ROIKVYVDNNNVYEES---LITLOKGDNDYQLLLSNLSN----REVFEEKIETINH 3931

OY 294 -----DMEE-----NIILSKNEDOSYONDSE-----FTTSKNTSTRHTSEVHG 335
 : : | : : : : : : : : : : : : : : : :
DB 3932 SNTNMFEDLEKLNGCVSNFFIOTKNTTYQMNDSSATIYGTFGVPNFKI-KSEDKILENN 3990

OY 336 AEV--HANSTSRTHSEVGHAENEHAVAIIDSLSLAGERTMAETMGANT--ADTAFL 389
 : : : : : : : : : : : : : : : : :
DB 3991 QOVVAMFPARKETIRDNTNWLOYTRPLKDYSTDFR-----EGTMAMDLSNSVNFEKEETT 4045

OY 390 NANIHYVTGTAPIYNVLEPTSILGKNQTLA-----TIKAKENOJSLAAPNNYTPSK 443
 |:|:|| : : : : : : : : : : : : : : : :
DB 4046 LVKITOFVKPKRAKNINNSENNAVILDONTSNSINSNYEFFTVGGHKLINTITSNNVNTNS 4105

OY 444 NLAPIALN-AODDESPPITMY--NOPLELEKTQLRDIDQYVGNIATYPFNCR--- 497
 : : : : : : : : : : : : : : : : :
DB 4106 OTINTLESIGVSKSWGKKIKILSYKSNDSISEIHINEVLIIESNKROYNILNLNRNTYT 4165

OY 498 ---VRDVTGSMSWEYLPOIQETTAIIFFNGKDLNVERRIAANVSPDET---K 547
 : : : : : : : : : : : : : : : : :
DB 4166 LIDVXLINNNVSDPRKCGNLNSTFITRTSAINVLANLEEISRNASTNLKTIKILIND 4225

OY 548 PDMFLKEA-----LKIAFGEHPENGNIQYOGKDIITEDDFPDFOOTSNIK-NOLAE 597
 ||| : : : : : : : : : : : : : : : :
DB 4226 PDNVLRDQQATTIYGNKKQAMGFITYSGNIKLYLATLVLDLFN-DKANVINISFNKP 4284

OY 598 LNATGYIVULKIKLNAAKMANNILRIDRPHVD-----RNNAIGA--DESVYEAARE 647
 : : : : : : : : : : : : : : : : :
DB 4285 IAAEN-----IGIDKSNMI-----YNDNISIKPLEINIIDIYNGPINKEYIVKANAOK 4332

OY 648 VINSTEGLLLNIDKDIRKILSGYIVELEDTEGLEKVID--RYMDMLNISL-RQDGTF 704
 - - - - - : : : : : : : : : : : : : : : :
DB 4333 --NNIDVDLGQIINKIAHNLR-FIAKFKFSIN-----NDIIFTNIVINGSSLVNNDGTS 4383

OY 705 IDFKKTYNDKPLYISNPAWK-VNYIYAATKENT 735
 ||| : : : : : : : : : : : : : : : :
DB 4384 IRFTLINMK-----ANKLYSLDVDYYLVNNS 4410

RESULT 6
AB6827
hypothetical protein yqfC [imported] - Lactococcus lactis subsp. lactis (strain IL140
C)Species: Lactococcus lactis subsp. lactis
CDate: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
CAccession: AB6827
R:Biolotin, A.; Wincker, P.; Mauger, S.; Jallou, O.; Malarme, K.; Weissenbach, J.; Eh
genome Res. In press, 2001
A>Title: The complete genome sequence of the lactic acid bacterium.

A:Reference number: A86625
A:Accession: A86827
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1072 <STO>
A:Cross-references: GI:AB005176; NID:q12724625; PIDN:AAK05715.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yq1c

Query Match 5.0%; Score 187; DB 2; Length 1072;
Best Local Similarity 18.8%; Pred. No. 0.043;
Matches 146; Conservative 151; Mismatches 333; Indels 148; Gaps 29;

```
QY 7 LIPALSTIIIVSSTGNLEIVQAEVQKQENRLNESESSQGLLGYFSDNLFQAPMVYTS 66
DB 319 LYPISAEASVMDNTLNTSLSSDSSIS-----SQTENSSQSG-----A 355
QY 67 STTGDLSTPSSLENIPEENQYFQSAISGFLTKVKSDEYTFATADNHVTMMVDDQEVY 126
DB 356 SSTAIISYDSENSLSNSQ-----INSNSSEKDSNQSSLSGSSMSSESEHSNSNI 410
QY 127 NKASNSNKR-----LEKGRLYQIKIYQORENPEKGLDFKL-----YWTDSQNKKEVTS 177
DB 411 NETNNSSEITNLTPSPNTPESNSVSDQTSSEASTNSNSISLSPSNISSTSDSESATNS 470
QY 178 DNLQLELKQKSSNSRKRKSTSGAPTPVDRNDGI-----PDSLEVEGYTVDKKRTFLS 233
DB 471 DFNVAEVAENNSLAVNSSSSVLSSTADNLGINQSDSLTKRD-----SSEISTSGAFLS 528
QY 234 PMISNIHKKGLTKYKS-----SPEKWS-----TASDPYDFEKVYGRIDKNVSPKAR 281
DB 529 ---SNQTSSEASTNSNSISLSPSNISSTVLESTTSSNSFNVAVANNSLAVNSSS 585
QY 282 HPLVAAYPIVHVDM-----ENILISKNEQO-----STQNTSETRTISKNTSRTHTS 330
DB 586 SVLSSTSTADNLEINQFSDNITKDSSEISTSGAFLSNQTSSEASNSMSINSPLSL 645
QY 331 EYHGAENVHANTSTSRHTSEYHGAENVHAVAIDHSLSLAGERTWAETGTLTADTARLN 390
DB 646 SLTNSSESATNOSNS--SEATKYDNNSSTHSSNILNS-----GSDSDSDSDSDSDSSNL- 699
QY 391 ANIRVNTGTAPIYVNLPTSLVGLKNOTLATIKAKENQLO-----ILAPNYYISKMLA 446
DB 700 -----SSPNLETTNOTIISKPEVNNISNPCKVSSNSVQENSTDH 741
QY 447 PIALNAQDFESTPTTMVNOFLELEKTQQLRDTQVYGINATYN---FENGRRVAVDTG 503
DB 742 EMSTPKKSISPISTSSQKESQSN---LLNTTEGINNPITNNSSSEKASASILT- 797
QY 504 SNMSEVLPQIOETTARIIF-----NGKDL-----NLVERRIAVNSDPLETTKPD 550
DB 798 -SYSNNSESSSTGLYISNEAQRNGSEISHLSPSSNSNENNVSSISQALLESSEKST 856
QY 551 TLK--PALKIAGFENP--NGNLQYQKDTTEFPNFDOOSNIKQMLAELNATNTYIVD 608
DB 857 NKRSSSLIINSTHPQEDNDQNSD-----EVKSNVNESIILQGLNLSIKKTHN 908
QY 609 KI---KLAKNNILIRDRFRHYDRNNIIVAGADESVVKEAHREIVNSSTEGLLNLDKDR 665
DB 909 SLTSGKLKVITLPSKSVYTNKKNNSMNVSEKLIKTPQKN--DSSQNLGQITALDLSFN 967
QY 666 KILSGYIEIEDTGLKEVIN-----DRYDMLNSSLROD--GKTFIDFKKYYDKL 714
DB 968 K-----EVTMEDSKTVPDKVLNDENGDRSQNNKSTIADKKKVFPEKRFSEFSKI 1018
```

RESULT 7
553805
alpha-toxin - Clostridium novyi (ATCC 19402)
C:Species: Clostridium novyi
A:Variety: ATCC 19402

C:Date: 28-Oct-1996 #sequence_revision 08-Nov-1996 #text_change 05-Nov-1999
C:Accession: S55805; S71294; S71158; S44273; I40834; S44272
C:Holmann, F.; Herrmann, A.; Habermann, E.; von Eichel-Streiber, C.
DOI: Gen. Genet. 247, 670-679, 1995
A:Title: Sequencing and analysis of the gene encoding the alpha-toxin of Clostridium
A:Reference number: I40834; MUID:95342160
A:Accession: S55805
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-2178 <HOF>
A:Cross-references: EMBL:Z48636; NID:g728537; PIDN:CAA8565.1; PID:g755724
A:Accession: S71294
A:Molecule type: protein
A:Residues: 1-15 <HOW>
R:Holmann, F.
Submitted to the EMBL Data Library, March 1995
A:Reference number: S71158
A:Accession: S71158
A:Molecule type: DNA
A:Residues: 1-1179, 'LKV', 'LWIGE', '1191-2178 <HOS>
A:Cross-references: EMBL:Z48636; NID:g728537; PIDN:CAA8565.1; PID:g755724
R:Holmann, F.; Habermann, E.; von Eichel-Streiber, C.
Submitted to the EMBL Data Library, July 1993
A:Description: Sequence analysis of Clostridium novyi alpha-toxin: a member of the fa
A:Reference number: S44272
A:Accession: S44273
A:Molecule type: DNA
A:Residues: 1-243; 1204-2178 <HON>
A:Cross-references: EMBL:Z23281
C:Genetics:
A:Gene: tcn-alpha
C:Superfamily: cpl repeat homology
C:Keywords: virulence factor
E:1880-1899/Domain: cpl repeat homology <CP12>

Query Match 4.9%; Score 183.5; DB 2; Length 2178;
Best Local Similarity 20.6%; Pred. No. 0.18; Mismatches 241; Indels 235; Gaps 34;

```
QY 122 DOEVINKASNSKIRLEKGRLYQI-KIYQOREN-----PTEKGLDFKL----- 163
DB 202 DDIINFLSNFK-----YDICKLNQKNNKNNKIAIGATQINENILTNKLSGY 253
QY 164 YWTDSQNKKEVTSNLOLPELKQKSSNSRKRKSTSGAPTPVDRNDGIPDSE-VEGYT 222
DB 254 YOELIOTNNLAAASDILRIALILKKYG-----GVYCDLDFLPQVN 292
QY 223 VDVKRKRTFLSPWISN-----IHEKKGLTKYKSSPEKWSFASDPYSDFEKVGR 272
DB 293 LSLFNDISKPNQMDSNYWEALFEALIANDEKMLNNY-----PYKIMEQVPS 340
QY 273 DKNVSPKAR---HPLVAAYPIVHVDMENILSKNEQOSTONTSETRTISKNTS-TSRT 327
DB 341 KERILSFVRNHINDILPLDGKIKISOLEILSLRKAATGKKTFSNAPLISNDSLFLNN 400
QY 328 HTSEVHG-----NAVYHANTSTSRHTSEYHGAENVHAVAIDHSLSLAGERTWAETG-L 381
DB 401 LISOLENREYELINSIOEKFKICETFYDSVINSSELVETTPKKNLSMDGSSFYQOIIGYL 460
QY 382 NTADTARNANIRY-----VNTGTAPIYVNLPTT--SLVIGKNOTLATIKAKENQLOSLA 435
DB 461 SSGFPEVNSYVFFSGPNITYSATCDTYHFKINTDMSSQOEI-----FEA 508
QY 436 PNNYPSKRLAPIALNAQDFESSPTI---TMVNOFLELEKTQQLRDTQVYGINATYN 492
DB 509 SNMLYFSK-----TDFEFKSWLNSNIAEKFOQLIKITYIGR-----TLN 549
QY 493 FENGR-----VRVDTGSNMSEVLPQIOETTARIIFNGKQNLVERRIAVNSDPLETTK 547
DB 550 YEDGILNFKKMKRVTT-----SELLKYLIEVNSPKIYENDLNNI-----LQIQG 593
QY 548 PDMTLKEALKIAFGFNERNGNLQYQKDTTEFPNFDQ--QTSQNIKQMLAELN----- 599
```


Db 1300 LKDMKVPDVVYSRSE---IPDNILVPPNTITFKKRGITAKRIVCGDQTSPPDTYSVI 1356
QY 354 GNAEVHAAVAIDHSLSLAGEERT-WAETMGILNTADT-ARLNANIRVYNTGTAPIYVNLV-7T 410
Db 1357 TTESLNHNHIIKIFLMIANNRNMFMKTLIDINHAFLYAKLEEE-----IYIPHPDR 1406
QY 411 SLVLCNQTATIKKKNQ-----LSQI-LAPNNYF-----SKNLAPIALNAQDD 455
Db 1407 RCVVNLNALVGLKQSPKEMWMDHLROYLNGIGLKDQNSYTPGLYQTEDKNLM-IAVYVDDC 1465
QY 456 FSSPTITMNYNOFL-ELEKTKOLR-----LPTDQVYGNIAIYNGENGAV----- 498
Db 1466 VTAASNEQRLDEFINKLSNFELKITGTLIDVLDFTD-ILGMDLYVNMKRLGITDILKSF 1524
QY 499 --RVDTSNMSEVLPQIOETTARIITFNGKDLNVERRIAANVP-SDPLETTKPD-----M 550
Db 1525 INRMD--KKYNEELKIKRKSIPHMSTYK-----IDPKDQVLOMSEEPERQGV 1571
QY 551 TLKELKALAFGENEPNGNLQYOGKDIETEPDFPDQOTSNIKQALAEIATNIIYVLDKI 610
Db 1572 KIQQLL-----GELNTYV-RHKCRDIEFAVKVAKRLVNPHERVYMIYKIQYL 1620
QY 611 KLNAAKNILIRDKRHYDRN-----NIAVGADESIVVKEAHR-----VI----- 649
Db 1621 -----VRKDIGIHYDRDCKKDKVAITPDASVGEYDQASRGVILMIGMIFNYYS 1673
QY 650 NSSTEGLLNIDKDIRKILSGY-----IYIEEDT-----EGIK--- 682
Db 1674 NKSTNRCSVSTFEALHAIYEGADSEFLKVLKELGEGDNNDIVMTDSKPAIQGLNSY 1733
QY 683 -----EVINDRYDMNITSSLRQDGKTF-----DRKKY 710
Db 1734 QOPKEFTWIKETIIEKIKRKSILKLTGKGNITADLTTRKVSASDEKRF 1784

RESULT 10
T18351
lmp1 protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 07-Dec-1999
C:Accession: T18351
R:Jensen, L.T.; Ladefoged, S.; Birkelund, S.; Christiansen, G.
Infected. Immun. 63, 3336-3347, 1995
A:Title: Selection of Mycoplasma hominis Pg21 deletion mutants by cultivation in the pre
A:Reference number: Z18884; MUID:95369882
A:Accession: T18351
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Cross-references: EMBL:U21961; NID:g790241; PID:g790242; PIDN:AAA81012.1
C:Genetics:
A:Gene: lmp1
A:Genetic code: SGC3

Query Match 4.8%; Score 179.5; DB 2; Length 1051;
Best Local Similarity 19.1%; Pred. No. 0.1;
Matches 152; Conservative 123; Mismatches 324; Indels 195; Gaps 30;

QY 19 SSTGNLEVIAQVQENLNLSESSSOGILGYFSDLNFOAPMVYTS---TTGDSLTP 75
Db 320 SNKSDIESANTELKQALAKANADKQVADNLAKEISKEQLNNSVSNANTLSAKLTDXDNTIQ 379
QY 76 SSELNIPSENOYFOSAIWSGFIKKKSDERTFATSADNHVTMWVDDQEVINKASNS--N 133
Db 380 QAKTE-LEKEIQKANQATIKSNNTJASMSQAS-----SSLDAAKVAEITRKLETFKDKDEAFN 434
QY 134 KIRLEKGLVQIKIOYQENPTEKGL-----DFKLYWDSQMKKEVISSDNIQLPEL 185
Db 435 ELKQIRNOIOEF-INTNKNPNYSSELLISQLTNSKRDSKNSVTDSSNKSDIESANT---EL 489
QY 186 KOKSSNSKKKSTASGPIVPPDRDNGIPDSLEVEGYTVDVKNKRFLSPWISNHE----- 241

Db 490 KOALNTAKAKSSI-----DNELRLKNDLQSKIEEFGIRINTNFSWISSKLETTKN 541
QY 242 --KKGILTK---YKSSPEKWSFA-SDPYSDFEKVTRIDKNVSP-----ARH 282
Db 542 KLAEEITLADAKIKNNPSSSKQALKDSSQOYKLGNELKLTIEEGKVEYETKNSNGYRLF 601
QY 283 PLVAAVPYIVHVDENI-----ILSKNEDQSTQNTDSETRITSKNTSTSRHTSEVH 333
Db 602 KLAQAEQFNNSDVIDKLNAWEKQTLISKQKLGQNSTKDYLTQLSTEMSTOESTIRKVI 661
QY 334 GNAEVHANTSTSRHTSEVHGAEVHAAVAIDHSLSLAGEERTMAEWMG---INTADTA-- 388
Db 662 VNIQAHIRNNLSQYRLPADKLIAMKRGYGDGVIESIQKWQDMDDSDSVSDSLKDD 721
QY 389 LMANIR-----YVNTGTAPIYVNLPTTSLVLCNQTATIRAKE----- 427
Db 722 FPKALRVLVGDTKPNPVSSWFINRNSIEYQULRLILVREHEIILLDAKAKDLKRAE 781
QY 428 -----NOLSOILAPNNYPSKNLAPIALNAQDDFSSPTITMNYNOFLEKTKOL 477
Db 782 KTIKFDENINSIDQ-----RAKRILQOELINKAKNDLSNFTLNHQKNOFTAKDITPKI 833
QY 478 RUDTQGVGNATYVFEKGRVAVDTSNMSEVLPQIOETT---ARIIFNGKDLNVERR 533
Db 834 SL-----LENKLEIN-----OYLLPIIKRAVASKISEIKRNRKLEDIIRS 875
QY 534 IAAVNPDPLETTKPDMLKELKALAFGENEPNGNLQYOGKDIETEPDFPDQOTSNIK 593
Db 876 NFYLMEKVEINKIYSELINKQY-----ELRSMINEK-----WESTKD 914
QY 594 QLAELNATNIIYVLDKIKLNAKM--NILIRDKRHYDRNNTIAVGADESIVVEAH----- 645
Db 915 TLDNIN-----IKENVSILKEVIINNSNAQYSINRILSTVPE-FIKVAQOTTRSN 963
QY 646 -REVINSSTEGLLNIDKDIRKILSGYIYIEEDTFGELKEVINDRYDMNITSSLRQDGKTF 704
Db 964 LRSIAEIOFKPPLIDIDKNKEV-----KKILDEN-----KTLNSD----- 999
QY 705 IDFKRYNDKRLPLYI 718
Db 1000 -NIAKLNEKIRLVL 1012

RESULT 11
T30822
lmp1 protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 07-Dec-1999
C:Accession: T30822
R:Jensen, L.T.; Ladefoged, S.; Birkelund, S.; Christiansen, G.
Infected. Immun. 63, 3336-3347, 1995
A:Title: Selection of Mycoplasma hominis Pg21 deletion mutants by cultivation in the
A:Reference number: Z18884; MUID:95369882
A:Accession: T30822
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1365 <JEN>
A:Cross-references: EMBL:U21962; NID:g790243; PID:g790244; PIDN:AAA81013.1
C:Genetics:
A:Gene: lmp1
A:Genetic code: SGC3

Query Match 4.8%; Score 179.5; DB 2; Length 1365;
Best Local Similarity 19.1%; Pred. No. 0.15;
Matches 152; Conservative 123; Mismatches 324; Indels 195; Gaps 30;

QY 19 SSTGNLEVIAQVQENLNLSESSSOGILGYFSDLNFOAPMVYTS---TTGDSLTP 75
Db 634 SNKSDIESANTELKQALAKANADKQVADNLAKEISKEQLNNSVSNANTLSAKLTDXDNTIQ 693
QY 76 SSELNIPSENOYFOSAIWSGFIKKKSDERTFATSADNHVTMWVDDQEVINKASNS--N 133

Db 694 OAKTE-LEKEIOKANOAIKSNMTASMOQSAK-----SSLDAAKVAEITKLETFNFKKEAKFN 748
QY 134 KIRBLEGRILYOIKIOYORKNPTREKGL-----DFKLWTOSQNKKEVYSSNLOLDEL 185
Db 749 ELKQTRNQIOER-INTKNNPNVSELISQLTSKRDSKNSVTSOSSNDSIESANT-----EL 803
QY 186 KOKSSNRKRKSTASAPVTPDRDNDGIPDSEVEGYTVDVKNKRTFLSPMISNIHE-----241
Db 804 KQALNTAKAKKSSI-----DNELRPLKNDLOSKIEEFGPLRANPNFMSISKLETTKN 855
QY 242 --KKGITK--YKSSPEKWSA--SDPYDSEKVTGRIDKNVSE-----ARH 282
Db 856 KLAELTELKADAIKNPSSKQALKDSSQOVOKLGNELLKTITFEFGKVEETKNSNIGYRL 915
QY 283 PLVAAYPIYHVDMENI-----ILSKNEDOSTONTSEFTTICKNTSTSTHTSEVA 333
Db 916 KLAQABOFNNSOVDKLKNAMEEKQTLTKKQKLGNOSTDYLTQJLSTEMSTOESTIKKYI 975
QY 334 GNAEYHANTSTSRHTSEVHGNAEVAHVAIDHSLSLAGERTVAETMG-----LNTADTAR-- 388
Db 976 VNIQAHIRNNLSQYRLKADKLIANKRGYGVKGIESIQKQMDLMDSDVSLVSDSLKD 1035
QY 389 LMANIR-----YVNTGTADPIYVNLPTTSLVLGKQTLATIKAKE-----427
Db 1036 FFKALRVLVGDYTKNPPVSWFINKRNRSEIENVQNLRLVRENEILLDKAKDLDRAE 1095
QY 428 -----NOLSOILANNNTYPSKNLAPIALNODDSSPTITANNQOFELETKOL 477
Db 1096 KTIKFEVDENINSLDQ-----RAKRLKOEILNANNDLSNFTLNHQKQFATAKDTTPKI 1147
QY 478 RLDTDOVYGNIAITYNENGVRVDTGSNMSEVLPJOIETT-----ARLIFNGKDLNIVER 533
Db 1148 SL-----LENLNLEIN-----QYLLPIIKERAVKISIELEKKNKLEEDIIKS 1189
QY 534 IAAVNSDPLETTKPDMLTEALKIAFGNEPNGNLQYOGKDTTEPDFNEDQOTSONIKN 593
Db 1190 NPYLMEKVEINIKYISELTNKQV-----ELRSNINFEK-----WESIKR 1228
QY 594 QLAELNATIIYVLDKIKILANAKM--NILRDKRFHIDRNNIAVGADESYYKEAH-----645
Db 1229 TLDNNI-----IKENVSILKEVIINNSNAQYSIRILSTVEP-FIKVAQTTRSN 1277
QY 646 -REVINSSTEGLLNIDKDIRKILSGYIEIEDTEGLKEIYNDRYOMLNISSRODGKTP 704
Db 1278 LKSLAEIOKRPFLDIDKMLKEY-----KILDEN-----KILNSD---- 1313
QY 705 IDFKRYNDKRLPYI 718
Db 1314 -NIAKLNEKIRLLV 1326
RESULT 12
PC1253
TYB protein - yeast (*Saccharomyces cerevisiae*) retrotransposon Ty4 (fragment)
N:Contants: integrase; proteinase; reverse transcriptase; RNase H
C:Species: *Saccharomyces cerevisiae*
C.Date: 30-Sep-1993 #sequence_rev150n 30-Sep-1993 #text_change 26-May-2000
C.Accession: PC1253; S27465
R:Stucka, R.; Schwarzlose, C.; Lochmuller, H.; Haecker, U.; Feldmann, H.
Gene 122, 119-128, 1992
A:Title: Molecular analysis of the yeast Ty4 element: Homology with Ty1, copia, and plan
A.Reference number: Jc1482; MUID:93083972
A.Accession: PC1253
A.Molecule type: DNA
A.Residues: 1-1467 <STU>
R:Stucka, R.; Schwarzlose, C.; Lochmuller, H.; Hcker, U.; Feldmann, H.
submitted to the EMBL Data Library, May 1992
A.Description: Molecular analysis of the yeast Ty4 element: homology with Ty1, copia, an
A.Reference number: S27465
A.Accession: S27465
A.Molecule type: DNA
A.Residues: 1-1467 <STW>
A.Cross-references: EMBL:M94164

C:Genetics:
A:Gene: TY4B
A:Mobile element: retrotransposon Ty4
C:Superfamily: TyB protein
Query Match 4.7%: Score 177.5; DB 2; Length 1467;
Best Local Similarity 19.7%: Pred. No. 0.21;
Matches 154; Conservative 121; Mismatches 277; Indels 231; Gaps 38;
QY 25 EVIOAEVYKQENRLNSESOGGLGYFSDLPFOAPMVYTSSTGDLSPSSLENIPS 84
Db 736 ELIQEDQKKTN--HETSPFEKESIG--TNVFRNNTNNEISLKTGDFSLPIKTLSEINN 789
QY 85 ENQYFOSAIWGCFLTKKSSDEYTFATSADNHVTMVDQEVINKASNKTRLEKGLYO 144
Db 790 HH-----SNDY-----STNKVE-----801
QY 145 IKIOYOREN-----PTEKGLDFKLYWTDQNKKEVYSSDNLOLPELAKOKSSNRKRKSTSA 200
Db 802 ---KFEKENHHPRIEDIVM-----SDQIDMESNCODGNLKEKLYTDKNAVPPDNGTNY 853
QY 201 GPTVPDRDNDGIPDSEVEGYTVDVKNKRTFLSPMISNIHEKK-----GITYX-- 248
Db 854 SPR-----LEQNIERSGSPVQTVNKSAPLNEKFSLSLMMKRRKRHRDKNSLTYELE 905
QY 249 -----KSSPEKWSASDPYSDPEKVTGRIDKNVSEPEARHPLVLAAYPIYHVDMEN 297
Db 906 RDKRSKRNKVKLIPDMETVASAKIRAIYVNEAISKNPDLKEHKEYKQAY--HKELON 962
QY 298 IILSKNEDOSTONDSE--TRTISKNTSRSHTSEVHGNAEYHANTSTSRHT-----349
Db 963 LKDMKVFVDVVKTSRSFIPDNLIVPTNTIFTKRNKGIYKARIYCRGCTGSPDITSVITTE 1022
QY 350 SEVHGNAEVAHVAIDHSLSLAGERTVAETMGTLTADT-ARLNMNIRVYVNTGTAPIYVNL 408
Db 1023 SLNHNHIIKILMOTTEICL-----W--TLDINHAFIYALKEETIYIPHLIGGVYKYL- 1074
QY 409 TTSVLVGNKQTLATIKAKENQ-----LSQI-LADNNYTP-----SKNLAPIALNAQ 453
Db 1075 -----NKALYGLKQSPKEMNDHLROYLNGIGLKONSYTGPGLYQTEDEKMLM-IAVYVD 1125
QY 454 DDESSPTITANNQOFL-ELKTKQLR-----LDTDOVYGNIAITYNENGVR-----498
Db 1126 DCVIAASNEQRLDEFIKLNSFELKITGTLIDVDLDT--ILGMDLVYNNRLGTLIDITLTK 1184
QY 499 ---RVDTGSNMSEVLPJOIETTARLIFNGKDLNLYERRIAAVNP--SDPLETTKPD---549
Db 1185 SFINRMD--KKYNEBELKIKKSSIPHMSTYK-----IDPKKDVLOMSEEPEROG 1231
QY 550 -MTLKEALKIAGFENGNGNLQYOGKDTTEPDFRFDQOTSONIKQLAELNATIIYVLD 608
Db 1232 VLKLOQL-----GELNYY-RHKCRYDIEFAKKAARLVNPHENFYVIYKIIO 1280
QY 609 KIKLAKMNLIDDKRFHYRN-----NIAVGADESYYKEAHRE-----VI-----649
Db 1281 YL-----VRKIDIGHYRDCKKDKKVAITDASGSEYDAQSRGIVLGMGMNFNV 1333
QY 650 --NSSTEGLLNIDKDIRKILSGYIEIEDTEGLKEIY-----NDRYDMLNIS--SLRQ 699
Db 1334 YSKNSTNRCSVSTAEALHAIYEGV---RDSFTLKVTLKELGSDNNNDIVMIVYKVTYKE 1389
QY 700 DGK 702
Db 1390 LGE 1392
RESULT 13
SS5101
hypothetical protein YMR219w - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: hypothetical protein YMR261.13; hypothetical protein YMR959.01
C:Species: *Saccharomyces cerevisiae*
C.Date: 08-Jul-1995 #sequence_rev150n 09-Mar-1996 #text_change 29-Oct-1999

Db 612 LGAISLINKENLSPITESEYEIPNDMLGLPLNDLNEKLEFNISKNTAYEFKKIYFNEDQ 671
Qy 410 -----TSLVLGNQTLA---TIKAKENQSLIAPNNYPSKRLPIALNA----- 452
Db 672 WMTQYQYQYFDLGMARSVLAQETLIRIKLOKLSYLIGSNISSDNLALMLTTNTL 731
Qy 453 QDDSPFPTMNN-----YNQFLELEKTQRLDLDQVYG--NIATYN 492
Db 732 RDISNEQIAMNNVDSFLNNAALCVFESNIPKISF-----MEQCINNINIKTE 782
Qy 493 FENGRAVDGGSNMSEVLPOIETTARIIFNGKDLNVERRIAANVSDPLETTKPMTL 552
Db 783 F-----IQKTNINDEKQLQNLN-----VENSLDFEFLN-----IQMKSLSESETALLI 829
Qy 553 KEAL-----KIAFGNENGN-----LOYQKDI-TEPPND-----QQTSON 590
Db 830 KEETWPELVLYAFKEGNNVIGDASGKNSTIEY-SKDIGLYVIGSDALYNGSNOSIS 888
Qy 591 IKNQLAELNATNTIYV-----LDKIKLNAMK----- 616
Db 889 FSNDFENGSLNFSSTIFWMLNKGDTIKSLKSGEDNGWEIYFODTGLVFNKIDSNG 948
Qy 617 ---NILRD---KRFHY-----DR--NNIAYGADESIV-KEAREVINSSTEGLLNLDK 662
Db 949 NEKNIYSDVSNNSMWHITTSVDRLEKQQLIFIDDLVANSIKEILNITSNIIISLSE 1008
Qy 663 DIRKILSGIYEIEDTGLKEVINDRYDMLNIISSLRDQKTFIDPK-----YNDKL 714
Db 1009 NNSYIGLITLKNPTTS-QEVLISNFEVLNNSYIRDSNEERLEYNKTYQLYNYVFSKDP 1067
Qy 715 PLYISNPNKYVNYAVATKENT 735
Db 1068 ICEVKQNN---NI-LTINN 1084

RESULT 15
S52611
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty4
N:Alternate names: protein YH008v-a
C:Species: Saccharomyces cerevisiae
C:Date: 05-May-1995 #sequence_revision 19-Oct-1995 #text_change 23-Mar-2001
C:Accession: S52611
R:Favell, T.
submitted to the EMBL Data Library, June 1994
A:Description: The sequence of S. cerevisiae cosmid L5018.
A:Reference number: S46798
A:Accession: S52611
A:Molecule type: DNA
A:Residues: 1-1802 <FAV>
A:Cross-references: EMBL:U01581
C:Genetics:
A:Map position: 8L
A:Mobile element: retrotransposon Ty4
C:Superfamily: Tyb protein

Query Match 4.68; Score 175; DB 2; Length 1802;
Best Local Similarity 20.1%; Pred. No. 0.38;
Matches 168; Conservative 118; Mismatches 276; Indels 274; Gaps 42;

Qy 25 EVIOAEVYKOEKRLNSESOGILGYFSDLNFOAPVAVVSSITGDLSPSELLENPS 84
Db 1072 ELIIO---EQFKTNHETSPKESIG---TNVFRNDNEISLKTGDTSPIKTLESINN 1125
Qy 85 ENQYFQSAIWSGFIKVKSDYEFTATSADNHVTMMVDDQEVINKASNSKIRLEKGRLYQ 144
Db 1126 HH-----SNDY-----STNKYE----- 1137
Qy 145 IKIIOYQEN---PTEKGLDFKLYWDSQNKKEYISSDNLQPLKQKSSNSRKKRSTSA 200
Db 1138 ---KFEKENHHPPIEDIVDM---SDQTMESNCQDGNMKELKLYDKNVPTDNGTNN 1189
Qy 201 GPVTPDRNDGIPDSLEVEGYTVDVKNKRTFLSPWISIHKK-----GLTKY--- 248

Db 1190 SPR-----LEQNIKESGSPVQTVNKSAPLNKFESSLNKKRRKRRDKNNSLSYLE 1241
Qy 249 -----KSSPEKSTASDPYSDEKYTG-----RIDKNVSPKARHPLVAAYPIVH 292
Db 1242 RDKRKRKRNRKVLIPDNETYS-----AQKIRAIYNEAISKNDLKEKHEYKOAY---H 1293
Qy 293 VDMENILSKNEDOSTQNTDSETRISKN---TSQRTSTSEVHG--MEVIANSTSRTH 348
Db 1294 KELQNLKMKYFDVQKTSRSE---IPDNLVPTWITFKRNGIYKARIYCRGDQSPD 1350
Qy 349 TSEVHGAIEVAVAIDHSLSLAGERT-WAETGLTADT-ARLANIRIYVVTGAPITNV 406
Db 1351 TYSVITPESLNHNHKKIFLMIANNRNMFKLIDINHALYAKLEEE-----IYIP 1400
Qy 407 LP-TTSLVLGNQTLATIKAKENQ-----LSOI-LAPNNYV-----SKNLAPIAL 450
Db 1401 HPHDRRCVYKLNKALYGLKQSPKEWMDHLROYLNGIGLKDNSYFGLYQTEDEKNIM-IAY 1459
Qy 451 NAQDDFSTPTMNNQFL-ELEKTKOLR-----LDDQVYGNATYNEFNGRY-- 498
Db 1460 YVDCVIAASNEQRLDEFINKLSNFEKLTGTLDVLDYD-ILGMDLYNKRRLGTIDL 1518
Qy 499 -----RVDTGSNMSEVLPOIETTARIIFNGKDLNVERRIAANV-SDPLETTKPD- 549
Db 1519 TLKSFINRMD--KYNELKRIKRSIPHMSTYK-----IDPKDYLOMSEEF 1565
Qy 550 ---MTLEALKIAFGNEPNGNLQYQKDIITEFDNFEDQOTSCNIKQLAELNATNIYT 605
Db 1566 RGVLKTLQQL-----GELNYV-RHKCRDYINFAVKVARLVNYPHERVYMYIK 1614
Qy 606 VLDRKILAKNANILIRDKRFHYDRN---NIAYGADESIVKEARE-----YI----- 649
Db 1615 ITQYLV-----VRYKDIGHYDRDCNKKVIALITDASVSEYDAQSRIGVILWYGMNI 1667
Qy 650 -----NSSTEGLLNIDIRKILSGY-----IYEIEDT-----EG 680
Db 1668 FNVYSNKTNCVSTSEALAIYGYADSETIKATLKELEGENDIYMTIDSPALQ 1727
Qy 681 LK-----EVINDRYDMLNIISSLRDQKTFI-----DEKVV 710
Db 1728 LNRSYQPKREKFTWIKTEIKIKESIKILKITKGKNIADLLTKPVASADPKRF 1783

Search completed: December 2, 2001, 13:51:36
Job time: 331 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:49:34 ; Search time 55.64 Seconds
(without alignments)
297.267 Million cell updates/sec

Title: US-09-747-521-4_COPY_1_735
Perfect score: 3765
Sequence: 1 MKRRKVLPLMALSTILVSS.....LYISNPYKVNYATKENT 735

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3519	93.5	735	1 US-08-021-601-4	Sequence 4, Appl1
2	3519	93.5	735	1 US-08-082-849B-4	Sequence 4, Appl1
3	3519	93.5	735	1 PCT-US94-01624-4	Sequence 4, Appl1
4	3519	93.5	903	1 US-08-021-601-12	Sequence 12, Appl1
5	3519	93.5	903	1 US-08-082-849B-12	Sequence 12, Appl1
6	3519	93.5	903	5 PCT-US94-01624-12	Sequence 12, Appl1
7	3342	88.8	719	1 US-08-082-849B-31	Sequence 31, Appl1
8	3342	88.8	719	1 PCT-US94-01624-31	Sequence 31, Appl1
9	783.5	20.8	884	2 US-08-471-033-5	Sequence 5, Appl1
10	783.5	20.8	884	2 US-08-471-044-5	Sequence 5, Appl1
11	783.5	20.8	884	2 US-08-463-483A-5	Sequence 5, Appl1
12	783.5	20.8	884	2 US-08-471-046A-5	Sequence 5, Appl1
13	783.5	20.8	884	2 US-08-470-566B-5	Sequence 5, Appl1
14	783.5	20.8	884	2 US-08-469-334-5	Sequence 5, Appl1
15	783.5	20.8	884	3 US-09-300-529-5	Sequence 5, Appl1
16	783.5	20.8	1346	1 US-08-471-033-23	Sequence 23, Appl1
17	783.5	20.8	1346	2 US-08-471-044-23	Sequence 23, Appl1
18	783.5	20.8	1346	2 US-08-463-483A-23	Sequence 23, Appl1
19	783.5	20.8	1346	2 US-08-471-046A-23	Sequence 23, Appl1
20	783.5	20.8	1346	2 US-08-470-566B-23	Sequence 23, Appl1
21	783.5	20.8	1346	2 US-08-469-334-23	Sequence 23, Appl1
22	783.5	20.8	1346	3 US-09-300-529-23	Sequence 23, Appl1
23	781	20.7	881	4 US-08-960-780-32	Sequence 32, Appl1
24	781	20.7	881	4 US-09-073-898-32	Sequence 32, Appl1
25	776.5	20.6	784	4 US-09-371-913A-7	Sequence 7, Appl1
26	771	20.5	1338	1 US-08-471-033-50	Sequence 50, Appl1
27	771	20.5	1338	2 US-08-471-044-50	Sequence 50, Appl1

28	771	20.5	1338	2 US-08-463-483A-50	Sequence 50, Appl1
29	771	20.5	1338	2 US-08-471-046A-50	Sequence 50, Appl1
30	771	20.5	1338	2 US-08-470-566B-50	Sequence 50, Appl1
31	771	20.5	1338	2 US-08-469-334-50	Sequence 50, Appl1
32	771	20.5	1338	3 US-09-300-529-50	Sequence 50, Appl1
33	769	20.4	852	1 US-08-471-033-36	Sequence 36, Appl1
34	769	20.4	852	2 US-08-471-044-36	Sequence 36, Appl1
35	769	20.4	852	2 US-08-463-483A-36	Sequence 36, Appl1
36	769	20.4	852	2 US-08-471-046A-36	Sequence 36, Appl1
37	769	20.4	852	2 US-08-470-566B-36	Sequence 36, Appl1
38	769	20.4	852	2 US-08-469-334-36	Sequence 36, Appl1
39	769	20.4	852	3 US-09-300-529-36	Sequence 36, Appl1
40	755.5	20.1	834	1 US-08-471-033-21	Sequence 21, Appl1
41	755.5	20.1	834	2 US-08-471-044-21	Sequence 21, Appl1
42	755.5	20.1	834	2 US-08-463-483A-21	Sequence 21, Appl1
43	755.5	20.1	834	2 US-08-471-046A-21	Sequence 21, Appl1
44	755.5	20.1	834	2 US-08-470-566B-21	Sequence 21, Appl1
45	755.5	20.1	834	2 US-08-469-334-21	Sequence 21, Appl1

ALIGNMENTS

RESULT 1
US-08-021-601-4
Sequence 4, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Kimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Aroca, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414,057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
FAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-4

Query Match 93.5% Score 3519, DB 1, Length 735;
Best Local Similarity 97.5% Pred. No. 7.1e-254;
Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;
QY 30 EVKQENRLNLSSESSQCLIGYFSDLNFGAPMVTSTGDISPSELENIPSENGYF 89

```
Db 1 EVKQENRLNSESSESSQGLGTYFSDNLOQAPMVYTSSTTGDLSIPSSLENIIPSENOYF 60
Qy 90 QSAIWSGFIKVKKSDYEYFATSADNHVTMWDQEVINKASNSNKRILEKRGLYOIKIOY 149
Db 61 QSAIWSGFIKVKKSDYEYFATSADNHVTMWDQEVINKASNSNKRILEKRGLYOIKIOY 120
Qy 150 QRENTEPEKLDKLTWDSQNKKEVYSSDNLQPELKOKSSNRKRKRSSTAGPYPPDRN 209
Db 121 QRENTEPEKLDKLTWDSQNKKEVYSSDNLQPELKOKSSNRKRKRSSTAGPYPPDRN 180
Qy 210 DQIPDSLEVEGYTVGVKKNRTFLSPWISNIHEKKGLTKKSSPEKMSSTASDYSDEKVT 269
Db 181 DQIPDSLEVEGYTVGVKKNRTFLSPWISNIHEKKGLTKKSSPEKMSSTASDYSDEKVT 240
Qy 270 GRIDKNVSPEARHPLVAAPYIVHDMENIILSKNEDOSTONTDSETRTSKNTSRHT 329
Db 241 GRIDKNVSPEARHPLVAAPYIVHDMENIILSKNEDOSTONTDSETRTSKNTSRHT 300
Qy 330 SEVHGNAEYHANTSRRTTSEVHGNAEYHAVAIDHSLSLAERTWAETMGLNTADTARL 389
Db 301 SEVHGNAEYHANTSRRTTSEVHGNAEYHAVAIDHSLSLAERTWAETMGLNTADTARL 360
Qy 390 NANIRYVNTGTAPIYVNLPTTSLVGLKNQTLATIKAKENQLSQILAPNNYYSKNIAPLA 449
Db 361 NANIRYVNTGTAPIYVNLPTTSLVGLKNQTLATIKAKENQLSQILAPNNYYSKNIAPLA 420
Qy 450 LNAODDFSSPTITMNYNOFLELEKTKQLRLDQYGYGNATYVNFENGRAVYDGSNMSEV 509
Db 421 LNAODDFSSPTITMNYNOFLELEKTKQLRLDQYGYGNATYVNFENGRAVYDGSNMSEV 480
Qy 510 LPQIOTETARIIIFNGKDLNVERRIAANVSPDLETTKPDMLKALKIAFGFNEPENGUL 569
Db 481 LPQIOTETARIIIFNGKDLNVERRIAANVSPDLETTKPDMLKALKIAFGFNEPENGUL 540
Qy 570 QYOGKDITEFDENFPOQTSQNIKNQLAELNATNITYVLDKIKLNKMNILINDKRFHYDR 629
Db 541 QYOGKDITEFDENFPOQTSQNIKNQLAELNATNITYVLDKIKLNKMNILINDKRFHYDR 600
Qy 630 NNIAVGADESIVKKAHREYINSTEGLLINDKDIRKILSGYIVETEDTEGLKEYINDRY 689
Db 601 NNIAVGADESIVKKAHREYINSTEGLLINDKDIRKILSGYIVETEDTEGLKEYINDRY 660
Qy 661 DMLNTSSLRQDQKTFIDFKKYNDKLPYISNPYKVVNYAVYKENT 706
Db 661 DMLNTSSLRQDQKTFIDFKKYNDKLPYISNPYKVVNYAVYKENT 706

RESULT 2
US-08-082-849B-4
: Sequence 4, Application US/08082849B
: Patent No. 5677274
: GENERAL INFORMATION:
: APPLICANT: Lepola, Stephen H.
: APPLICANT: Klompel, Kurt R.
: APPLICANT: Arota, Naveen
: APPLICANT: Singh, Yogendra
: APPLICANT: Nichols, Peter J.
: TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
: NUMBER OF SEQUENCES: 35
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/082, 849B
: FILING DATE: 25-JUN-1993
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/021,601
: FILING DATE: 12-FEB-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Weber, Kenneth A.
: REGISTRATION NUMBER: 31,677
: REFERENCE/DOCKET NUMBER: 15280-161-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 735 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-082-849B-4
```

```
Query Match 93.5%: Score 3519; DB 1; Length 735;
Best Local Similarity 97.5%: Pred. No. 7,1e-234;
Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;
```

```
Qy 30 EVKQENRLNSESSESSQGLGTYFSDNLOQAPMVYTSSTTGDLSIPSSLENIIPSENOYF 89
Db 1 EVKQENRLNSESSESSQGLGTYFSDNLOQAPMVYTSSTTGDLSIPSSLENIIPSENOYF 60
Qy 90 QSAIWSGFIKVKKSDYEYFATSADNHVTMWDQEVINKASNSNKRILEKRGLYOIKIOY 149
Db 61 QSAIWSGFIKVKKSDYEYFATSADNHVTMWDQEVINKASNSNKRILEKRGLYOIKIOY 120
Qy 150 QRENTEPEKLDKLTWDSQNKKEVYSSDNLQPELKOKSSNRKRKRSSTAGPYPPDRN 209
Db 121 QRENTEPEKLDKLTWDSQNKKEVYSSDNLQPELKOKSSNRKRKRSSTAGPYPPDRN 180
Qy 210 DQIPDSLEVEGYTVGVKKNRTFLSPWISNIHEKKGLTKKSSPEKMSSTASDYSDEKVT 269
Db 181 DQIPDSLEVEGYTVGVKKNRTFLSPWISNIHEKKGLTKKSSPEKMSSTASDYSDEKVT 240
Qy 270 GRIDKNVSPEARHPLVAAPYIVHDMENIILSKNEDOSTONTDSETRTSKNTSRHT 329
Db 241 GRIDKNVSPEARHPLVAAPYIVHDMENIILSKNEDOSTONTDSETRTSKNTSRHT 300
Qy 330 SEVHGNAEYHANTSRRTTSEVHGNAEYHAVAIDHSLSLAERTWAETMGLNTADTARL 389
Db 301 SEVHGNAEYHANTSRRTTSEVHGNAEYHAVAIDHSLSLAERTWAETMGLNTADTARL 360
Qy 390 NANIRYVNTGTAPIYVNLPTTSLVGLKNQTLATIKAKENQLSQILAPNNYYSKNIAPLA 449
Db 361 NANIRYVNTGTAPIYVNLPTTSLVGLKNQTLATIKAKENQLSQILAPNNYYSKNIAPLA 420
Qy 450 LNAODDFSSPTITMNYNOFLELEKTKQLRLDQYGYGNATYVNFENGRAVYDGSNMSEV 509
Db 421 LNAODDFSSPTITMNYNOFLELEKTKQLRLDQYGYGNATYVNFENGRAVYDGSNMSEV 480
Qy 510 LPQIOTETARIIIFNGKDLNVERRIAANVSPDLETTKPDMLKALKIAFGFNEPENGUL 569
Db 481 LPQIOTETARIIIFNGKDLNVERRIAANVSPDLETTKPDMLKALKIAFGFNEPENGUL 540
Qy 570 QYOGKDITEFDENFPOQTSQNIKNQLAELNATNITYVLDKIKLNKMNILINDKRFHYDR 629
Db 541 QYOGKDITEFDENFPOQTSQNIKNQLAELNATNITYVLDKIKLNKMNILINDKRFHYDR 600
Qy 630 NNIAVGADESIVKKAHREYINSTEGLLINDKDIRKILSGYIVETEDTEGLKEYINDRY 689
Db 601 NNIAVGADESIVKKAHREYINSTEGLLINDKDIRKILSGYIVETEDTEGLKEYINDRY 660
Qy 690 DMLNTSSLRQDQKTFIDFKKYNDKLPYISNPYKVVNYAVYKENT 735
Db 690 DMLNTSSLRQDQKTFIDFKKYNDKLPYISNPYKVVNYAVYKENT 735
```

Db 661 DMLNSSLRODGKTFIDFKKYNKDLPLYISNPYKVVAVTKENT 706

RESULT 3

PCT-US94-01624-4

Sequence 4, Application PC/TUS9401624

GENERAL INFORMATION:

APPLICANT: Leppia, Stephen H.

APPLICANT: Kilmpel, Kurt R.

APPLICANT: Atora, Naveen

APPLICANT: Singh, Yogendra

APPLICANT: Nichols, Peter J.

TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND

TITLE OF INVENTION: RELATED METHODS

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESSES:

ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW

STREET: Stewart Street Tower, 20th Floor, One Market

STREET: Plaza

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94105

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/01624

FILING DATE: June 25, 1993

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Weber, Kenneth A.

REGISTRATION NUMBER: 31,677

REFERENCE/DOCKET NUMBER: 15280-115

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 543-9600

TELEFAX: (415) 543-5043

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 735 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US94-01624-4

Query Match 93.5%; Score 3519; DB 5; Length 735;

Best Local Similarity 97.5%; Pred. No. 7, 1e-254;

Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 30 EYKQENRLNSESSESSQGLIGYFSDLNFOAPNVVTSSTTGDLSIPSSLELNIPSENOYF 89

Db 1 EYKQENRLNSESSESSQGLIGYFSDLNFOAPNVVTSSTTGDLSIPSSLELNIPSENOYF 60

QY 90 QSAIWSGFTIKVKSDEYTPATSDNHNVTMVDQEVYINKASNSNKRIRLENGRLYQIY 149

Db 61 QSAIWSGFTIKVKSDEYTPATSDNHNVTMVDQEVYINKASNSNKRIRLENGRLYQIY 120

QY 150 QRENPEKGLDFKLYTDSQNKKEVTSNDLQJPELKOKSSNRKRKSTAGPTVPDRDN 209

Db 121 QRENPEKGLDFKLYTDSQNKKEVTSNDLQJPELKOKSSNRKRKSTAGPTVPDRDN 180

QY 210 DGIPLDSLEVEGYVDVKNKRTFLSPMISNIHEKKGILTKYSSPEKWTASDPYSDPEKVT 269

Db 181 DGIPLDSLEVEGYVDVKNKRTFLSPMISNIHEKKGILTKYSSPEKWTASDPYSDPEKVT 240

QY 270 GRIDKAVSPEARHPPLVAAPYIVVDMENITLSKNEDOSTONTSETRTISKNTSTSTHT 329

Db 241 GRIDKAVSPEARHPPLVAAPYIVVDMENITLSKNEDOSTONTSETRTISKNTSTSTHT 300

QY 330 SEVHGNAEYHANTSTRTHTSEVHGNAEYHAVAIDHSLSLAGERTMAETGLNTADTARL 389

Db 301 SEVHGNAEYHASFEDIGGSVSAGFSNNSSTVAIDHSLSLAGERTMAETGLNTADTARL 360

QY 390 NANIRYVNTGTAPLYNLPPTSLVLGKNOGLATIKAKENOLSOILAPNNYPSKNLAPIA 449

Db 361 NANIRYVNTGTAPLYNLPPTSLVLGKNOGLATIKAKENOLSOILAPNNYPSKNLAPIA 420

QY 450 LNAODFSSPTITMNNVQFLEKTKQLRLDTQVYVNIATYFENGRVAVDQGSNNSEV 509

Db 421 LNAODFSSPTITMNNVQFLEKTKQLRLDTQVYVNIATYFENGRVAVDQGSNNSEV 480

QY 510 LPQIETARTIIFNGKDLNVERRIAIVNSDPLETKPMTLEALKIIFGENEPGNL 569

Db 481 LPQIETARTIIFNGKDLNVERRIAIVNSDPLETKPMTLEALKIIFGENEPGNL 540

QY 570 QYQKDTIEFDENFDQTSQNIKNQLAELNATNYYVLDKIKLNKKNLIRDRKFHYDR 629

Db 541 QYQKDTIEFDENFDQTSQNIKNQLAELNATNYYVLDKIKLNKKNLIRDRKFHYDR 600

QY 630 NNIAVGADSVKAEHREVINNSTEGILLNIDKIRKILSGYIEIEDTEGLKEVINDRY 689

Db 601 NNIAVGADSVKAEHREVINNSTEGILLNIDKIRKILSGYIEIEDTEGLKEVINDRY 660

QY 690 DMLNSSLRODGKTFIDFKKYNKDLPLYISNPYKVVAVTKENT 735

Db 661 DMLNSSLRODGKTFIDFKKYNKDLPLYISNPYKVVAVTKENT 706

RESULT 4

US-08-021-601-12

Sequence 12, Application US/08021601

Patent No. 5591631

GENERAL INFORMATION:

APPLICANT: Leppia, Stephen H.

APPLICANT: Kilmpel, Kurt R.

APPLICANT: Nichols, Peter J.

APPLICANT: Atora, Naveen

APPLICANT: Singh, Yogendra

TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND

TITLE OF INVENTION: RELATED METHODS

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Needle & Rosenberg, P.C.

STREET: 133 Carnegie Way, Suite 400

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/021,601

FILING DATE: 19930212

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Spratl, Gwendolyn D.

REGISTRATION NUMBER: 36,016

REFERENCE/DOCKET NUMBER: 1414,057

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404/688-0770

TELEFAX: 404/688-9880

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 903 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-021-601-12

Query Match 93.5%; Score 3519; DB 1; Length 903;
Best Local Similarity 97.5%; Pred. No. 9.7e-254;
Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

30 EVKQENRLNESESSOGGLGYFSDLNFOAPMVYTSSTGDLSPSSLENIPEENQYF 89
1 EVKQENRLNESESSOGGLGYFSDLNFOAPMVYTSSTGDLSPSSLENIPEENQYF 60
90 OSAISGFLTKVKSDEYFPATISADNHVYMWVDQEVINKASNSNRIREKGRLYOIKIY 149
61 OSAISGFLTKVKSDEYFPATISADNHVYMWVDQEVINKASNSNRIREKGRLYOIKIY 120
150 QRENTEGGLDEKLTWYTSQNKKEVYSSDNLOLPELKOKSSNRKRSTASAPYPPDRN 209
121 QRENTEGGLDEKLTWYTSQNKKEVYSSDNLOLPELKOKSSNRKRSTASAPYPPDRN 180
210 DGIPLSLEVEGYTVYVKNKRFLSPWISNIEHKGLTKYKSSPEKMSASDPYSDFEKYT 269
181 DGIPLSLEVEGYTVYVKNKRFLSPWISNIEHKGLTKYKSSPEKMSASDPYSDFEKYT 240
270 GRIDKNVSPERHPLVAAYPIYHVMENIILSKNDOSTQNTDSTRTISKNTSRHT 329
241 GRIDKNVSPERHPLVAAYPIYHVMENIILSKNDOSTQNTDSTRTISKNTSRHT 300
330 SEVHGNAEYHANTSRHTSEVHGNAEYHAYADHSLSLAGEERTWAGTGLNTADTARL 389
301 SEVHGNAEYHANTSRHTSEVHGNAEYHAYADHSLSLAGEERTWAGTGLNTADTARL 360
390 NANIRYVNTGTAPIYVNLPTSLVIGKNQTLATIKAKENQSLIAPNNYPSKMLADIA 449
361 NANIRYVNTGTAPIYVNLPTSLVIGKNQTLATIKAKENQSLIAPNNYPSKMLADIA 420
450 LNAODDESSPTITMANYNOFLELEKTKOURLTDQYGYGNATYINFENGVRVDGSMSEV 509
421 LNAODDESSPTITMANYNOFLELEKTKOURLTDQYGYGNATYINFENGVRVDGSMSEV 480
510 LPOIETARIIFNCKDLNVERRIAAYNPSPLETTKPDMTLKALIAFGFNPNGNL 569
481 LPOIETARIIFNCKDLNVERRIAAYNPSPLETTKPDMTLKALIAFGFNPNGNL 540
570 QYQGDITEFEPNPDQOSTONIKNOELAEINATNTYVLDKIKLNKMMILLRDKRFHNDY 629
541 QYQGDITEFEPNPDQOSTONIKNOELAEINATNTYVLDKIKLNKMMILLRDKRFHNDY 600
630 NNIAVGADESVYKKAHREYINSSTEGLLNIDKDIRKILSGYIEIEDTEGLKEVINDRY 689
601 NNIAVGADESVYKKAHREYINSSTEGLLNIDKDIRKILSGYIEIEDTEGLKEVINDRY 660
690 DMLNSSLRQDGKTFIDPKKYNDKLPYISNPYKVNYAYATKENT 735
661 DMLNSSLRQDGKTFIDPKKYNDKLPYISNPYKVNYAYATKENT 706

RESULT 5
US-08-082-849B-12
Sequence 12, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Kimpel, Kurt R.
APPLICANT: Arora, Navveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-12

Query Match 93.5%; Score 3519; DB 1; Length 903;
Best Local Similarity 97.5%; Pred. No. 9.7e-254;
Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

30 EVKQENRLNESESSOGGLGYFSDLNFOAPMVYTSSTGDLSPSSLENIPEENQYF 89
1 EVKQENRLNESESSOGGLGYFSDLNFOAPMVYTSSTGDLSPSSLENIPEENQYF 60
90 OSAISGFLTKVKSDEYFPATISADNHVYMWVDQEVINKASNSNRIREKGRLYOIKIY 149
61 OSAISGFLTKVKSDEYFPATISADNHVYMWVDQEVINKASNSNRIREKGRLYOIKIY 120
150 QRENTEGGLDEKLTWYTSQNKKEVYSSDNLOLPELKOKSSNRKRSTASAPYPPDRN 209
121 QRENTEGGLDEKLTWYTSQNKKEVYSSDNLOLPELKOKSSNRKRSTASAPYPPDRN 180
210 DGIPLSLEVEGYTVYVKNKRFLSPWISNIEHKGLTKYKSSPEKMSASDPYSDFEKYT 269
181 DGIPLSLEVEGYTVYVKNKRFLSPWISNIEHKGLTKYKSSPEKMSASDPYSDFEKYT 240
270 GRIDKNVSPERHPLVAAYPIYHVMENIILSKNDOSTQNTDSTRTISKNTSRHT 329
241 GRIDKNVSPERHPLVAAYPIYHVMENIILSKNDOSTQNTDSTRTISKNTSRHT 300
330 SEVHGNAEYHANTSRHTSEVHGNAEYHAYADHSLSLAGEERTWAGTGLNTADTARL 389
301 SEVHGNAEYHANTSRHTSEVHGNAEYHAYADHSLSLAGEERTWAGTGLNTADTARL 360
390 NANIRYVNTGTAPIYVNLPTSLVIGKNQTLATIKAKENQSLIAPNNYPSKMLADIA 449
361 NANIRYVNTGTAPIYVNLPTSLVIGKNQTLATIKAKENQSLIAPNNYPSKMLADIA 420
450 LNAODDESSPTITMANYNOFLELEKTKOURLTDQYGYGNATYINFENGVRVDGSMSEV 509
421 LNAODDESSPTITMANYNOFLELEKTKOURLTDQYGYGNATYINFENGVRVDGSMSEV 480
510 LPOIETARIIFNCKDLNVERRIAAYNPSPLETTKPDMTLKALIAFGFNPNGNL 569
481 LPOIETARIIFNCKDLNVERRIAAYNPSPLETTKPDMTLKALIAFGFNPNGNL 540
570 QYQGDITEFEPNPDQOSTONIKNOELAEINATNTYVLDKIKLNKMMILLRDKRFHNDY 629
541 QYQGDITEFEPNPDQOSTONIKNOELAEINATNTYVLDKIKLNKMMILLRDKRFHNDY 600
630 NNIAVGADESVYKKAHREYINSSTEGLLNIDKDIRKILSGYIEIEDTEGLKEVINDRY 689

Db 601 NNIAVGADESVEKAEHREVINSSTEGLLNIDKIRKILSGYIEVEDTGLKEVINDRY 660
QY 690 DMLNISSLRDQKTFIDFKKRYNDKPLIYSNPYKVVAVTKENT 735
Db 661 DMLNISSLRDQKTFIDFKKRYNDKPLIYSNPYKVVAVTKENT 706

RESULT 6

PCT-US94-01624-12
Sequence 12, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
STREET: Stewart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ. ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-12

Query Match 93.5%; Score 3519; DB 5; Length 903;
Best Local Similarity 97.5%; Pred. No. 9, 7e-254;
Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 30 EYVQENRLNESSESSGGLIGYFSDINFOAPMVYSSSTGGDLSIPSELENIPSENOYF 89
Db 1 EYVQENRLNESSESSGGLIGYFSDINFOAPMVYSSSTGGDLSIPSELENIPSENOYF 60
QY 90 QSAIWSGFTVKKSDETFTATSADNHVTMWVDDOEVIKNSNSKTRLEKGRLYQIQY 149
Db 61 QSAIWSGFTVKKSDETFTATSADNHVTMWVDDOEVIKNSNSKTRLEKGRLYQIQY 120
QY 150 QRENPTKGLDFKLYWTDSONKKEVSSDNQLPELKOKSSNSRKRSTAGPTVPDRDN 209
Db 121 QRENPTKGLDFKLYWTDSONKKEVSSDNQLPELKOKSSNSRKRSTAGPTVPDRDN 180
QY 210 DGIPLDSLEVEGYVVDVKNKFTFLSPWISNIHEKKGLTKRYSSPEKWSASDPYSDFEYVT 269
Db 181 DGIPLDSLEVEGYVVDVKNKFTFLSPWISNIHEKKGLTKRYSSPEKWSASDPYSDFEYVT 240

QY 270 GRIDKNVSPARHPPLVAAPYIVHVDMENIILSKNEDSTONTDEFTTISKNTSRTHT 329
Db 241 GRIDKNVSPARHPPLVAAPYIVHVDMENIILSKNEDSTONTDEFTTISKNTSRTHT 300
QY 330 SEVHGAHNAEVAHNTSTSTHTSEVHGAHNAEVAHVAIDHSLSLAGERTWMTGLTADTARL 389
Db 301 SEVHGAHNAEVAHNTSTSTHTSEVHGAHNAEVAHVAIDHSLSLAGERTWMTGLTADTARL 360
QY 390 NAIIRYVNTGTAPLYNLPFTSLVLGNQTLATIKAKENOLSLIAPNNYPPSKNLAPIA 449
Db 361 NAIIRYVNTGTAPLYNLPFTSLVLGNQTLATIKAKENOLSLIAPNNYPPSKNLAPIA 420
QY 450 LNAODDESSPTITMNYNQFLEKTKQLRDTDOVGNITATYFENGRVAVDGSNNSEV 509
Db 421 LNAODDESSPTITMNYNQFLEKTKQLRDTDOVGNITATYFENGRVAVDGSNNSEV 480
QY 510 LPOIETTARIIFNGKDLNVERRIAANVPSDPLETKPMTLEKALIAFGFNEPNCNL 569
Db 481 LPOIETTARIIFNGKDLNVERRIAANVPSDPLETKPMTLEKALIAFGFNEPNCNL 540
QY 570 QYQKDTTEFDNFNDQOTSONIKNQLAELNATNITYVLKIKLAKANNILIRDKRFHYDR 629
Db 541 QYQKDTTEFDNFNDQOTSONIKNQLAELNATNITYVLKIKLAKANNILIRDKRFHYDR 600
QY 630 NNIAVGADESVEKAEHREVINSSTEGLLNIDKIRKILSGYIEVEDTGLKEVINDRY 689
Db 601 NNIAVGADESVEKAEHREVINSSTEGLLNIDKIRKILSGYIEVEDTGLKEVINDRY 660
QY 690 DMLNISSLRDQKTFIDFKKRYNDKPLIYSNPYKVVAVTKENT 735
Db 661 DMLNISSLRDQKTFIDFKKRYNDKPLIYSNPYKVVAVTKENT 706

RESULT 7

US-08-082-849B-31
Sequence 31, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ. ID NO: 31:

SEQUENCE CHARACTERISTICS:
LENGTH: 719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
us-08-082-849B-31

Query Match 88.8%; Score 3342; DB 1; Length 719;
Best Local Similarity 92.8%; Pred. No. 1e-240;
Matches 659; Conservative 6; Mismatches 21; Indels 24; Gaps 2;

```
OY 30 EYKQENRLINESSSOGGLGYFSDNLNQAPMVYTSSTGDLSTPSSLENIPESENOYF 89
    |||||||
DB 1 EYKQENRLINESSSOGGLGYFSDNLNQAPMVYTSSTGDLSTPSSLENIPESENOYF 60
OY 90 OSAISGFIKVKKSDYFATPSADNHVPMWVDDQEVINKASNSNRILEKGLYOIKIOY 149
    |||||||
DB 61 OSAISGFIKVKKSDYFATPSADNHVPMWVDDQEVINKASNSNRILEKGLYOIKIOY 120
OY 150 QRENPTKGLDFKLYWTDSONKKEVISSDNQLPELKOKSSNTATIMQKGFLOGPTVP 205
    |||||||
DB 121 QRENPTKGLDFKLYWTDSONKKEVISSDNQLPELKOKSSNTATIMQKGFLOGPTVP 180
OY 206 DRDNGCIDPSLEVEGYTVDVKNKRTFLSPWISNIHEKKGLTKYKSSPEKWSASDPYSDF 265
    |||||||
DB 181 DRDNGCIDPSLEVEGYTVDVKNKRTFLSPWISNIHEKKGLTKYKSSPEKWSASDPYSDF 240
OY 266 EKVYGRIDKNVSPKARHPLVAAYPIVHVDMEIILSKNEDOSTONTDSETRISKNTSTS 325
    |||||||
DB 241 EKVYGRIDKNVSPKARHPLVAAYPIVHVDMEIILSKNEDOSTONTDSETRISKNTSTS 300
OY 326 RTHTSEVHGNAEVAHANTSTSRTHTSEVHGNAEVAHVAIDHSLSLAGERTWAEIATGLNTAD 385
    |||||||
DB 301 RTHTSEVHGNAEVAHANTSTSRTHTSEVHGNAEVAHVAIDHSLSLAGERTWAEIATGLNTAD 360
OY 386 TARLANANIRYVNTGAPLIYNVLPPTTSLVIGKNOTLATITKAKENQSLIAPANNYPYSKNL 445
    |||||||
DB 361 TARLANANIRYVNTGAPLIYNVLPPTTSLVIGKNOTLATITKAKENQSLIAPANNYPYSKNL 420
OY 446 APIALNAODDFSSPTITMNYNOFLEKTKQLRLDTPDYVGNIAIYTNENGVRVDTGSN 505
    |||||||
DB 421 APIALNAODDFSSPTITMNYNOFLEKTKQLRLDTPDYVGNIAIYTNENGVRVDTGSN 460
OY 506 WSEVLPQIOETTARLIIFNGKDLNLVERRIAAYVNSDPLETTKPDMTLKEALKIAGFNEP 565
    |||||||
DB 461 WSEVLPQIOETTARLIIFNGKDLNLVERRIAAYVNSDPLETTKPDMTLKEALKIAGFNEP 520
OY 566 NGNLOYGKDIETEPFNFDQOOTSNIKNOIAELNATNTIYTVLDIKLNAKMMILIRDRKF 625
    |||||||
DB 521 NGNLOYGKDIETEPFNFDQOOTSNIKNOIAELNATNTIYTVLDIKLNAKMMILIRDRKF 580
OY 626 HYDRNNIAVGADESYYKKAHREVIINSTEGLNLIDKIDRIKLSGIYIEIDTEGLKEVI 685
    |||||||
DB 581 HYDRNNIAVGADESYYKKAHREVIINSTEGLNLIDKIDRIKLSGIYIEIDTEGLKEVI 640
OY 686 NDRIDMLNISSLRDQKTFIDFKKYNKLIPLYISNPNKRVNVAVTKEPT 735
    |||||||
DB 641 NDRIDMLNISSLRDQKTFIDFKKYNKLIPLYISNPNKRVNVAVTKEPT 690
```

RESULT 8
PCT-US94-01624-31
Sequence 31, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppala, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arota, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAKX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESS: TOWNSEND and TOWNSEND KHOURIE and CREW
STREET: Stewart Street Tower, 20th Floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043

INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:
LENGTH: 719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-31

Query Match 88.8%; Score 3342; DB 5; Length 719;
Best Local Similarity 92.8%; Pred. No. 1e-240;
Matches 659; Conservative 6; Mismatches 21; Indels 24; Gaps 2;

```
OY 30 EYKQENRLINESSSOGGLGYFSDNLNQAPMVYTSSTGDLSTPSSLENIPESENOYF 89
    |||||||
DB 1 EYKQENRLINESSSOGGLGYFSDNLNQAPMVYTSSTGDLSTPSSLENIPESENOYF 60
OY 90 OSAISGFIKVKKSDYFATPSADNHVPMWVDDQEVINKASNSNRILEKGLYOIKIOY 149
    |||||||
DB 61 OSAISGFIKVKKSDYFATPSADNHVPMWVDDQEVINKASNSNRILEKGLYOIKIOY 120
OY 150 QRENPTKGLDFKLYWTDSONKKEVISSDNQLPELKOKSSNTATIMQKGFLOGPTVP 205
    |||||||
DB 121 QRENPTKGLDFKLYWTDSONKKEVISSDNQLPELKOKSSNTATIMQKGFLOGPTVP 180
OY 206 DRDNGCIDPSLEVEGYTVDVKNKRTFLSPWISNIHEKKGLTKYKSSPEKWSASDPYSDF 265
    |||||||
DB 181 DRDNGCIDPSLEVEGYTVDVKNKRTFLSPWISNIHEKKGLTKYKSSPEKWSASDPYSDF 240
OY 266 EKVYGRIDKNVSPKARHPLVAAYPIVHVDMEIILSKNEDOSTONTDSETRISKNTSTS 325
    |||||||
DB 241 EKVYGRIDKNVSPKARHPLVAAYPIVHVDMEIILSKNEDOSTONTDSETRISKNTSTS 300
OY 326 RTHTSEVHGNAEVAHANTSTSRTHTSEVHGNAEVAHVAIDHSLSLAGERTWAEIATGLNTAD 385
    |||||||
DB 301 RTHTSEVHGNAEVAHANTSTSRTHTSEVHGNAEVAHVAIDHSLSLAGERTWAEIATGLNTAD 360
OY 386 TARLANANIRYVNTGAPLIYNVLPPTTSLVIGKNOTLATITKAKENQSLIAPANNYPYSKNL 445
    |||||||
DB 361 TARLANANIRYVNTGAPLIYNVLPPTTSLVIGKNOTLATITKAKENQSLIAPANNYPYSKNL 420
OY 446 APIALNAODDFSSPTITMNYNOFLEKTKQLRLDTPDYVGNIAIYTNENGVRVDTGSN 505
    |||||||
DB 421 APIALNAODDFSSPTITMNYNOFLEKTKQLRLDTPDYVGNIAIYTNENGVRVDTGSN 460
OY 506 WSEVLPQIOETTARLIIFNGKDLNLVERRIAAYVNSDPLETTKPDMTLKEALKIAGFNEP 565
    |||||||
DB 461 WSEVLPQIOETTARLIIFNGKDLNLVERRIAAYVNSDPLETTKPDMTLKEALKIAGFNEP 520
```



```

1 NUMBER OF SEQUENCES: 50
2 CORRESPONDENCE ADDRESS:
3 ADDRESSEE: CIBA-GEIGY Corporation
4 STREET: 7 Skyline Drive
5 City: Hawthorne
6 STATE: NY
7 COUNTRY: USA
8 ZIP: 10532
9
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Floppy disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: Patentin Release #1.0, Version #1.30B
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/471,044
17 FILING DATE: 06-JUN-1995
18 CLASSIFICATION: 800
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: US 08/463,483
21 FILING DATE: 05-JUN-1995
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: US 08/314,594
24 FILING DATE: 09-SEP-1994
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: US 08/218,018
27 FILING DATE: 23-MAR-1994
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: US 08/037,057
30 FILING DATE: 25-MAR-1993
31 ATTORNEY/AGENT INFORMATION:
32 NAME: Pace, Gary M.
33 REGISTRATION NUMBER: 40,403
34 REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SQLv3
35 TELECOMMUNICATION INFORMATION:
36 TELEPHONE: 919-541-8582
37 TELEFAX: 919-541-8689
38 INFORMATION FOR SEQ ID NO: 5:
39 SEQUENCE CHARACTERISTICS:
40 LENGTH: 884 amino acids
41 TYPE: amino acid
42 TOPOLOGY: linear
43
44 MOLECULE TYPE: protein
45
46 US-08-471-044-5

```

[illegible]

Oy	331	EUNGAEEVNAHNTSPRSRTTSEVHNGAEVN-----AVAIHDSLSAGERTAEITG--	360
Db	334	---NSVEHSSTNNSTYTBEGASVDAIGSRKISRGVSANVOHSEYVAOE--WSTSTGNT	378
Oy	381	-LNTADTARLANIRIVNVTGTARIYVNLPTTSSLVLGKNQTLATIKKKNOLSPN	438
Db	379	SOENFASAGYLANIRVNRVNVGTGAIDYVKPRTTSPVL--NDNTIATPKASNSTALNISPE	437
Oy	439	YTPSKNLPALNADODSSPTPTMNTNQGLELEKTKQLRLDTQVYGNATYVFNENGRV	498
Db	438	SYPRKGQNGIAITSDDEDSNHPILNKKQVNNLNNRPMLETPETQOTG---VYKIKDPTG	494
Oy	499	RVDTSNMNSEVLPLOEFTTATIRNGKDLIVERIRIAVNPNSDLETTKPDMLTKLKI	558
Db	495	NIVTGGEMNGVYIOQIKATTAISLYDQGE-RAEKRVAAKDYENEDXT-PSLTIKDLAKL	552
Oy	559	AF--GENEPNGNLOYQGDITEPDF--NFDQOTSQNIKNQLAEL----NATNIYTVLQK	609
Db	553	SYPRDEIKIEGLLTVKKNPRIEYSSVMYTLIDENTAKETWKLQNDPTTGKFKFDVSHLYDV--	609
Oy	610	IKLANKNMILRLROKRFHNDNRNINIVGADSESVYKRAHREVNINSSREG-----LL	657
Db	610	-KLTPKKMAYVTK-LSILYDN---AESNDNSIGKATNTNINIVSGGNGNKKQYSSNNPDANLT	664
Oy	658	LNID-----KDIRKISGYIVELDETEGLEKVINDRYDMLNIS--SLRODGKFTFIDKKY	710
Db	665	LNTDQOEKLNNRDRYUISLYMKSEKNOCETITIGELIYPTTKTVNVNKNKDYKRLDIIAH	724
Oy	711	NDKLPLRYISNP 721	
Db	725	NIK-----SNP 730	

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```


FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Spruill, W. Murray
 REGISTRATION NUMBER: 32,943
 REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8615
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 884 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-463-483A-5

Query Match 20.8%; Score 783.5; DB 2; Length 884;
 Best Local Similarity 29.6%; Pred. No. 4,2e-50;
 Matches 234; Conservative 133; Mismatches 293; Indels 131; Gaps 29;

QY 1 MKRRKVLPLMALSTLVSS--TGNLEVIOAEVK-----QENRLINESESSOGLG 51
 DB 1 MKMKKKLASVVTCTLLAPMFLNGNVAAYADSKTNQISTQKNQ---QKEMRKGLG 57
 QY 52 YFSDLNFOAPMVTSTGDLSPSELEN--IPSENQFQSAIWSGFIYKKSDEYTF 109
 DB 58 YFKGKDF-SULTMFAPTRDSTLIYDOOTANKLDDKQOEYQSTIRWIGLISKETGDTFN 116
 QY 110 TSADNHVTMVDQEVINNAASNSKIRLEKGRLYQIRIYQRENPTEKGLD-----FKL 163
 DB 117 LSEDEQALIEINCKIISNKEKQVHLEKGLVPIKIEYQSD--TKFNIDSKTFKELKL 174
 QY 164 YMTDSQNKKEVISDNULPELKQKS-----NSRKRKTSAGPIYPPDND 210
 DB 175 FKIDSONOPQVOODELRNEEFKKSQEFELAPSKINLFTQKKRREIDED--TDTDGD 231
 QY 211 GIPDSLEVEGYTVDKKRTFLSPWISNIHEKKGLTKYKSSPEKWSAPSPYDFEYVG 270
 DB 232 STFDLMEENKTYT-----QNIAYKMDSL-ASKGYTFEYVNPLESHTVGPDIYDEKAAR 286
 QY 271 RIDKNVSPEARHPLVAAYPIVHVMENIILSKNEDOSTONTDETRTISKNTSRTHTS 330
 DB 287 DLDLSNAKEFTFNPVLAFAFVSNVMEKVILSPNENIS----- 323
 QY 331 EVGNNAEVNANTSTSRHTSEVGNNAEVH-----AVALDLSLAGEETAETNG-- 380
 DB 324 ---NSVESHSSTWSTYNTGASVEAGIGPKGISFGVSVNYOHSEYVAQE--WGTSTGNT 378
 QY 381 --LNTADTALNINIRYVNTGTAPIYVNLPTTSLVGLKNOTLATITKAKENQLSILAPNN 438
 DB 379 SORNTASAGYLANVRNNVNGTAIYDKPTTSFVL--NNDITAITIKSSSTALNISPEGE 437
 QY 439 YPSKMLAPIALNAODSFSTPTTMNYNDFLELEKTQRLRLDQVYGNATYFENGGRV 498
 DB 438 SYPKKGONGAITSMDDENSHPTLKKKQVDNLNNKPMLEFNQDTG---VYKIKTHG 494
 QY 499 RVDTGSNMSVYLFOIQETTRITFNGKDLNVERRIAANVSPDLPTTKDMLTKELKI 558
 DB 495 NIVTGGEMNGVIOQIRAKTASIIIVDQGE-FVAERKVAAKYENPEDKT-PSLTLKDALKL 552
 QY 559 AF--GFNEPKNIOYOGKDTTEPDF--NPDQOTSQNIKNOLAEL-----NATNIYTVLDK 609
 DB 553 SYPDEIKETLGLLYKNKPIYESSVMTYLDENTAKETKQDLNDTGTGFKFVSHLYDV--- 609
 QY 610 IKLNAKMNILIRDKRPHYDRNNIYVGADESVEAHRVINSSTEG-----LL 657
 DB 610 -KLTPKMNVTIK-LSILYDN--AESDNSIGKWTNTNINISGNGKQKQYSSNNPDANLT 664
 QY 658 LNTD-----KDIRKILSGIYVEIDTEGLKEVINDRYDMLNIS-SLRDQKFTIDPKKY 710
 DB 665 LNTDAOEKLKMRNDYYISTLYMKSEKNTQCEITIDGIEIYPTTKTVNVNKKNDYKRLDIAH 724

QY 711 NDKLPLXISNP 721
 DB 725 NIK-----SNP 730

RESULT 12
 US-08-471-046A-5
 Sequence 5, Application US/08471046A
 Patent No. 5866326
 GENERAL INFORMATION:
 APPLICANT: Warren, Gregory W
 APPLICANT: Kozziel, Michael G
 APPLICANT: Mullins, Martha A
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desai, Nalin M
 APPLICANT: Kostichka, N. Kristy
 APPLICANT: Duck, Nicholas B
 APPLICANT: Estruch, Juan J
 TITLE OF INVENTION: Method for Isolating Vegetative Insecticidal
 TITLE OF INVENTION: Protein Genes
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 5866326artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: NC
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30B
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/471,046A
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/463,483
 FILING DATE: 05-JUN-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057
 FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 884 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-471-046A-5

Query Match 20.8%; Score 783.5; DB 2; Length 884;
 Best Local Similarity 29.6%; Pred. No. 4,2e-50;
 Matches 234; Conservative 133; Mismatches 293; Indels 131; Gaps 29;

```

QY 52 YFSDLNFOAPMVVTSSTGDLSPSESELEN--IPSENOYFQSAIWSGFIKVKKSDEYTF 109
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 YFKGKDF--SNLTMFAPTRDSTLIYDQGTANKLLDKKQOEGYSIRWIGLQSKETGDFTFN 116
QY 110 TSADNHVTMMVDDQEVINKASNSNKIRLEKGRLYQIKIOYQRENPTKGLD-----FKL 163
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 LSEDOAIIEINGKITISNGKEKQYVHLEKGLVPIKIEYQSD--TKFNIDSKTEKEKL 174
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 164 YWTDSONKKEVYSSDNLQPELKOKS-----NSRKRSTSAQTPVDROND 210
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 FRIDSONQPOVOQDELRLRPEFNKESQOFLAKPSKINLFTOKMREIDE--IDTQGD 231
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 211 GIPDSLEVEGYTVDVKNKRTFLSPWISNIHEKGLTKYKKSPEKSTASDPYSDEKYWG 270
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 SIPDLMEENGTYT-----QNRIVAKWDDSL-ASKGYTKFVSNPLESHYVDPYDYEKKAAR 286
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 271 RIDKAVSPRARPLVAAPPIVHVDENIILSKNEDOSTONTSETRTTSKNTSTSRTHS 330
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 287 DDLDSNAKETFPPLVAAPPSVAVSMKEKYLSPENILS----- 323
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 331 EVHGNAEVHANTSTSRTHSEVHNAEYH-----AVAIHSLSLAGERTWAETMG-- 380
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 334 ---NSVESHSTSNMSTYTNTEGASVAGIGPKGISFGVSNVOHSETVAOE--WGTSTGNT 378
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 381 --LNTADTARLANIRYVNTGTAPITYNLPTTSLVGLKNQTLATIKAKENQLSOLLAPNN 438
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 379 SQFNASAGYLNAANRYNVNGTGAIYDVPTTSPVL--NNDTIATITAKSNSFALNISPE 437
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 439 YPESKNLAPIALNADDDSSPTITNNYNOFLEKTKOLRLTDQVYGINATYNEENGAV 498
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 438 STPKGGQNGIATTSMDDESHPTITLNKKQVDLLNNKPMELTNOTDG--YVKIKDTHG 494
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 499 RVDTSNNSSEVLPJOETTARIIFNGKDLNVERRIAANVPSDPETTKPDMTLKEALKI 558
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 495 NIVTGEENNGVYQIKATKATASIIYVDGE--RVAKRVAAKDYENPDKT--PSLTLDAKL 552
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 559 AF--GFNEPNCMLYQOGKDTIEFDF--NFDQOTSONIKQALAE-----NATNIVTVLDK 609
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 553 STPDEIKETIEGLLYKKNPIYESSVMVTYLDENTAKEVTQOLDNTGKFDVSHLYDV-- 609
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 610 IKLNKMMILIRDKRFHNRDNNIAVGADESYYKHAHREYINSTSG-----LL 657
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 610 -KLTPKMTTK--LSILDN--ALSNDSTIGKMTNTNIVSGGNNKKOYSSNNDPANT 664
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 658 LNID-----KDIRKILSGYIIEDETEGLKEVINDRYDMLNIS--SLRODKTFIDFKKY 710
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 665 LNTDQGEKLNKRDYIISLYMKSEKNTQCEIITIDGEIYPTTKTYVNVKNDKRIIDIAH 724
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 711 NDKLPLYSNP 721
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 725 NIK-----SNP 730
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 13
 US-08-470-566B-5
 Sequence 5, Application US/08470566B
 Patent No. 5872212

GENERAL INFORMATION:

APPLICANT: Warren, Gregory M
 APPLICANT: Koziele, Michael G
 APPLICANT: Mullins, Martha A
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desai, Nalini M
 APPLICANT: Kostichka, N. Kristy
 APPLICANT: Duck, Nicholas B
 APPLICANT: Estruch, Juan J
 TITLE OF INVENTION: No. 5872212el pesticidal proteins and strains
 NUMBER OF SEQUENCES: 52
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 5872212artlis Corporation
 STREET: 3054 Cornwalis Road

```

CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/470,566B
  FILING DATE: 06-JUN-1995
  CLASSIFICATION: 530
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/463,483
    FILING DATE: 05-JUN-1995
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/314,594
    FILING DATE: 09-SEP-1994
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/218,018
    FILING DATE: 23-MAR-1994
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/037,057
    FILING DATE: 25-MAR-1993
  ATTORNEY/AGENT INFORMATION:
    NAME: Meigs, J. Timothy
    REGISTRATION NUMBER: 38,241
  REFERENCE/DOCKET NUMBER: CCG1695/CIP3/DIV4 - SOLV4
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 919-541-8587
    TELEFAX: 919-541-8689
  INFORMATION FOR SEQ ID NO: 5:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 864 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
    MOLECULE TYPE: protein
  US-08-470-566B-5

```

Query Match 20.8%; Score 783.5; DB 2; Length 864;
 Best Local Similarity 29.6%; Pred. No. 4.2e-50;
 Matches 234; Conservative 133; Mismatches 293; Indels 131; Gaps 29;

```

QY 1 MKRRVLLPIMLSTIIYSS--TGNLEVYQAEVK-----QENRLNSESSESGGLIGY 51
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MKNNKKLASVYTCITLAPMLNGVNAVYADSKTNQISTTKNO--QKENDRKGGLGY 57
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 52 YFSDLNFOAPMVVTSSTGDLSPSESELEN--IPSENOYFQSAIWSGFIKVKKSDEYTF 109
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 YFKGKDF--SNLTMFAPTRDSTLIYDQGTANKLLDKKQOEGYSIRWIGLQSKETGDFTFN 116
QY 110 TSADNHVTMMVDDQEVINKASNSNKIRLEKGRLYQIKIOYQRENPTKGLD-----FKL 163
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 LSEDOAIIEINGKITISNGKEKQYVHLEKGLVPIKIEYQSD--TKFNIDSKTEKEKL 174
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 164 YWTDSONKKEVYSSDNLQPELKOKS-----NSRKRSTSAQTPVDROND 210
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 FRIDSONQPOVOQDELRLRPEFNKESQOFLAKPSKINLFTOKMREIDE--IDTQGD 231
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 211 GIPDSLEVEGYTVDVKNKRTFLSPWISNIHEKGLTKYKKSPEKSTASDPYSDEKYWG 270
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 SIPDLMEENGTYT-----QNRIVAKWDDSL-ASKGYTKFVSNPLESHYVDPYDYEKKAAR 286
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 271 RIDKAVSPRARPLVAAPPIVHVDENIILSKNEDOSTONTSETRTTSKNTSTSRTHS 330
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 287 DDLDSNAKETFPPLVAAPPSVAVSMKEKYLSPENILS----- 323
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 331 EVHGNAEVHANTSTSRTHSEVHNAEYH-----AVAIHSLSLAGERTWAETMG-- 380
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 334 ---NSVESHSTSNMSTYTNTEGASVAGIGPKGISFGVSNVOHSETVAOE--WGTSTGNT 378

```


Db 725 NIK-----SNP 730

RESULT 15
US-09-300-529-5
Sequence 5, Application US/09300529
Patent No. 6066783
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalin M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: Genes Encoding Insecticidal Proteins
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6066783artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/300,529
FILING DATE: TBA
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/469,334
FILING DATE: 06-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-19506L
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-300-529-5

Query Match 20.8%; Score 783.5; DB 3; Length 884;
Best Local Similarity 29.6%; Pred. No. 4,2e-50;
Matches 234; Conservative 133; Mismatches 293; Indels 131; Gaps 29;

Qy 1 MKRKVLIPLMALSTIVS--TGNLEVIQAEK-----OENRLNSESSESSOGLG 51
Db 1 MKMKKLASVYTCILAPMLNGNNAVYADSKTNOISTOKNO---OKENDRKGLG 57

Qy 52 YPSDLNFOAPMVTSTTGLSIPSSLELN--IPSENQYFOASINSGFIKVKSDYEYFA 109
Db 58 YFKGKDF-SNLTMFAPTRSTLIIYDQGTANKKLDDKQGEYQSIWIGLQSKSETGDFTFN 116
Qy 110 TSADNHVTMVDQEVYINKASNSNKTREKGRLYQIKIOYQRENPTEKGLD-----FKL 163
Db 117 LSEDEQALIEINGKITISNKGKQOVVHLEKGLVPIKIEYQSD--TKFMDKTEKTELK 174
Qy 164 YWTDSONKKEVISDNLOLPELKOKSS-----NSRKRSTAGPTVPDROND 210
Db 175 FKIDSONPOOVQODELRNPEFNKESQEFLLAKPSKINLFTQMKREID--TDTQSD 231
Qy 211 GIPDSLEVEGYTVDKNKRFTLSPWISNIHERKGLTKYKSPKSTASDPYDEKYG 270
Db 232 SIPDLMEENGTYL----QNRIVKMDSL-ASKGYTKFVSNLSESTVGDPTDYEKAR 286
Qy 271 RIDKNVSPARHPVLAAYIVYVDMENITLSKNEODSTQNTDSETRTTSKNTSTSRHTS 330
Db 287 DDLISNAKETFPNLYAAPSVMVSMKVLSPNENLS----- 323
Qy 331 EVHGAENVHANTSTSRHTSEVHGAENVH-----AVADHSLSAGERTMAETMG-- 380
Db 324 ---NSVESHSTNMSTTNEGASVEAGICPKISRGVSVNYQHSYAOE--WGTSTGNT 378
Qy 381 --LNTADTARLANIRVYNTGTAPIYVNLPTTSLVLGKNOTLATIKAKENQLSQILAPNN 438
Db 379 SOFNTASAGYLANVYNNVGTGAIYDKPPTSFVL--NNDTJATTTAKSNSPALNISPE 437
Qy 439 YPSKRLAIALNADDDSEPTITMNYNOFLELEKTQRLDQDYGYAIATYNNENGRV 498
Db 438 STPKGQNGCIATISMDDFNSHPITLKKQVDNLNNKPMLETNOTDG--YKIKIDHG 494
Qy 499 RVDGSMSEVLPOIETTARIIFNCKDLNVERIAAVNPSDPLETTPKPDMLKEALKT 558
Db 495 NVTGGEWNGVIOQIKAKTASIIYDDGE--RVAEKRYAANDYENPEDKT--PSLTLDALK 552
Qy 559 AF--GFNEPNGILQYOGKDIIEFD--NPDQOTSONIKQLAEL-----NATNIYVLDK 609
Db 553 SVPEIKETIEGLLYKKNKPIYESSVMTYLDENTAKEVQNDJTGKFDVSHLDV--- 609
Qy 610 IKLNKMNILIDDKRHHYRNNAIANGADSVYKKAHREYINSTES-----LL 657
Db 610 -KLTPEKMYTIK-LSLILDN---AESDMSIGKMTNTIVSGNNGKQYSSNNDPANLT 664
Qy 658 LNID-----KDIRKILSGYIEIEDTEGLKEVINDRYMLNIS--SLRODGKTFIDPKY 710
Db 665 LMTDAQEKLNKRNDYIISLYMKSEKNTQCEIYIDGEIYITTKYVNVKNDYKRLDIAT 724
Qy 711 NDKLPLYISNP 721
Db 725 NIK-----SNP 730

Search completed: December 2, 2001, 13:49:43
Job time: 223 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:48:31 ; Search time 112.1 Seconds
(without alignments)
485.672 Million cell updates/sec

Title: US-09-747-521-4_COPY_1_735

Perfect score: 3765
Sequence: 1 MKRRVLLPLMALSTILVSS.....LYISNPKNVNYATKENT 735

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_1101: *
1: /SIDS8/gcgdata/geneseq/AA1980.DAT: *
2: /SIDS8/gcgdata/geneseq/AA1981.DAT: *
3: /SIDS8/gcgdata/geneseq/AA1982.DAT: *
4: /SIDS8/gcgdata/geneseq/AA1983.DAT: *
5: /SIDS8/gcgdata/geneseq/AA1984.DAT: *
6: /SIDS8/gcgdata/geneseq/AA1985.DAT: *
7: /SIDS8/gcgdata/geneseq/AA1986.DAT: *
8: /SIDS8/gcgdata/geneseq/AA1987.DAT: *
9: /SIDS8/gcgdata/geneseq/AA1988.DAT: *
10: /SIDS8/gcgdata/geneseq/AA1989.DAT: *
11: /SIDS8/gcgdata/geneseq/AA1990.DAT: *
12: /SIDS8/gcgdata/geneseq/AA1991.DAT: *
13: /SIDS8/gcgdata/geneseq/AA1992.DAT: *
14: /SIDS8/gcgdata/geneseq/AA1993.DAT: *
15: /SIDS8/gcgdata/geneseq/AA1994.DAT: *
16: /SIDS8/gcgdata/geneseq/AA1995.DAT: *
17: /SIDS8/gcgdata/geneseq/AA1996.DAT: *
18: /SIDS8/gcgdata/geneseq/AA1997.DAT: *
19: /SIDS8/gcgdata/geneseq/AA1998.DAT: *
20: /SIDS8/gcgdata/geneseq/AA1999.DAT: *
21: /SIDS8/gcgdata/geneseq/AA2000.DAT: *
22: /SIDS8/gcgdata/geneseq/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3652	97.0	764	AA56958	B. anthracis prote
2	3652	97.0	764	AA56958	Wild type B. anthr
3	3551.5	94.3	763	AA56960	B. anthracis TPA-P
4	3519	93.5	735	AA60179	Protective antigen
5	3519	93.5	736	AA56959	B. anthracis MAT-P
6	3519	93.5	903	AA60183	PA(1-725)-----Huma
7	3342	88.8	719	AA60193	Modified protectiv
8	2664	70.8	569	AA56961	B. anthracis PA63
9	787.5	20.9	884	AA91239	B. cereus VIP1A(a)
10	783.5	20.8	884	AA19509	B. cereus VIP1A(a)
11	783.5	20.8	884	AA46712	100 kDa VIP1A(a) t

12	783.5	20.8	1346	AA91245	VIP2A(a) and VIP1A
13	783.5	20.8	1346	AA19513	B. cereus VIP1A(a)
14	783.5	20.8	1346	AA46723	VIP1A(a)/VIP2A(a)
15	781.5	20.8	880	AA60224	Bacillus thuringie
16	781.5	20.7	881	AA59277	MIS toxin from B.
17	778.5	20.7	884	AA63793	Bacillus cereus 10
18	776.5	20.6	784	AA82948	MIS toxin of Bacil
19	771	20.5	1338	AA91247	VIP2A(a)-VIP1A(a)
20	771	20.5	1338	AA19520	Maize optimised-B.
21	771	20.5	1338	AA46731	VIP2A(a)/VIP1A(a)
22	769	20.4	852	AA91246	VIP1A(a) protein w
23	769	20.4	852	AA19516	Maize optimised-B.
24	769	20.4	852	AA46727	Maize optimised VI
25	766.5	20.4	860	AA59282	MIS-8 toxin from B
26	755.5	20.1	834	AA91242	B. thuringiensis V
27	755.5	20.1	834	AA19512	B. thuringiensis V
28	755.5	20.1	834	AA46722	Vegetative insect
29	596.5	15.8	425	AA60222	Bacillus thuringie
30	584	15.5	667	AA91240	B. cereus VIP1 pro
31	580	15.4	667	AA19510	B. cereus 80 kD VI
32	580	15.4	667	AA46713	80 kDa VIP1A(a) to
33	578	15.4	667	AA63794	Bacillus cereus 80
34	541	14.4	357	AA60228	Bacillus thuringie
35	538	14.3	357	AA60231	Bacillus thuringie
36	521	13.8	357	AA60218	Bacillus thuringie
37	521	13.8	357	AA60226	Bacillus thuringie
38	503	13.0	357	AA60227	Bacillus thuringie
39	450	12.4	327	AA60220	Bacillus thuringie
40	447	11.9	327	AA60223	Bacillus thuringie
41	427.5	11.4	348	AA60219	Bacillus thuringie
42	425.5	11.3	348	AA60229	Bacillus thuringie
43	419.5	11.1	347	AA60221	Bacillus thuringie
44	419.5	11.1	347	AA59276	Toxin from B. thur
45	396	10.5	333	AA60230	Bacillus thuringie

ALIGNMENTS

RESULT 1	
AA56958	AA56958 standard; protein: 764 AA.
XX	
XX	AA56958;
AC	
XX	
DT	25-APR-2000 (first entry)
XX	
DE	B. anthracis protective antigen (PA) protein.
XX	
KW	Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;
KW	tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.
XX	
OS	Bacillus anthracis.
XX	
PN	W0200002522-A2.
XX	
PD	20-JAN-2000.
XX	
PF	09-JUL-1999; 99WO-US15568.
XX	
PR	10-JUL-1998; 98US-0092416.
XX	
PA	(USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX	
PI	Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;
XX	
XX	WPI: 2000-182165/16.
DR	N-PSDB: AA256874.
XX	
XX	Recombinant DNA construct useful as vaccines for anthrax, in producing
PT	host cells for analyzing the drugs and agents inhibiting anthrax
XX	
PS	Disclosure: Page 33; 35pp; English.

XX The invention provides a recombinant DNA construct that comprises a
 CC vector and at least one nucleic acid (or its fragment) encoding a
 CC combination of Bacillus anthracis proteins, selected from protective
 CC antigen (PA), MAT-PA (PA with its secretory signals removed), rPA-PA
 CC (PA with its secretory signals replaced with those of tissue plasminogen
 CC activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
 CC for anthrax and in producing infectious alpha virus particles. These
 CC particles, expressing the B. anthracis proteins are useful also as
 CC vaccines for anthrax. Host cells transformed with the construct are
 CC useful for analyzing the effectiveness of drugs and agents that inhibit
 CC anthrax or B. anthracis proteins. The present sequence represents a
 CC B. anthracis PA protein.

XX Sequence 764 AA:

Query Match 97.0%; Score 3652; DB 21; Length 764;

Best Local Similarity 97.6%; Pred. No. 5.9e-237;

Matches 717; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

1 MKRRVLPIMLSTIVSSGNLEVIQAEVQENRLNESSSSOGILGYFSDLNFOA 60
 1 mkkrrvlpimlslstivssgnleviqaevgqenrlnessessgllgyfstdlnfqa 60
 61 PMVVTSSSTGDISPSSELENIPSENOYFQSAIWSGFIKKKSDEYTFATSADNHVTMV 120
 61 pmvvtstgdispsselelenipseenyfqsaiwsqfikkksdeytfatsadnhvtmv 120
 121 DDOEYINRASNKIRLEKGRLYQIKIQORENPTKGLDFLWYSDSNKKEVSSDL 180
 121 ddoeyinrASNKIRLEKGRlyqIKIQORENPTKGLDFLWYSDSNKKEVSSDL 180
 122 dQEVYINRASNKIRLEKGRlyqIKIQORENPTKGLDFLWYSDSNKKEVSSDL 180
 122 dQEVYINRASNKIRLEKGRlyqIKIQORENPTKGLDFLWYSDSNKKEVSSDL 180
 181 QLPPELKOKSSNRKRSTAGPTVDRDNDGIPDSLEVGYYVDVYKNTKTFSPMISNH 240
 181 qlpPElkOKSSNRKRSTAGPTVDRDNDGIPDSLEVGYYVDVYKNTKTFSPMISNH 240
 181 qlPElkOKSSNRKRSTAGPTVDRDNDGIPDSLEVGYYVDVYKNTKTFSPMISNH 240
 181 qlPElkOKSSNRKRSTAGPTVDRDNDGIPDSLEVGYYVDVYKNTKTFSPMISNH 240
 241 EKKGLTKTKSSPEKKSSTASDPYSDEKVTGRIDKNVSPARHPLVAAPYIVHVDENITL 300
 241 ekkGLTKTKSSPEKKSSTASDPYSDEKVTGRIDKNVSPARHPLVAAPYIVHVDENITL 300
 241 ekkGLTKTKSSPEKKSSTASDPYSDEKVTGRIDKNVSPARHPLVAAPYIVHVDENITL 300
 241 ekkGLTKTKSSPEKKSSTASDPYSDEKVTGRIDKNVSPARHPLVAAPYIVHVDENITL 300
 301 SKNEOSQNTDSEMTTSTKNTSTRTSEVHNAEVAANTSTKTHSEVHNAEVA 360
 301 skNEOSQNTDSEMTTSTKNTSTRTSEVHNAEVAANTSTKTHSEVHNAEVA 360
 301 skNEOSQNTDSEMTTSTKNTSTRTSEVHNAEVAANTSTKTHSEVHNAEVA 360
 301 skNEOSQNTDSEMTTSTKNTSTRTSEVHNAEVAANTSTKTHSEVHNAEVA 360
 361 VAIDHSLSLAGERTWAEETGLTADTARLANIRVYNGTAPYVNLPTTSLVLGKNQTL 420
 361 vaIDHSLSLAGERTWAEETGLTADTARLANIRVYNGTAPYVNLPTTSLVLGKNQTL 420
 361 vaIDHSLSLAGERTWAEETGLTADTARLANIRVYNGTAPYVNLPTTSLVLGKNQTL 420
 361 vaIDHSLSLAGERTWAEETGLTADTARLANIRVYNGTAPYVNLPTTSLVLGKNQTL 420
 421 ATIKAKENOLSIAPNNYPSKNIAPIALNAODESSPTTMMNTNOELEKTKQLRID 480
 421 atIKAKENOLSIAPNNYPSKNIAPIALNAODESSPTTMMNTNOELEKTKQLRID 480
 421 atIKAKENOLSIAPNNYPSKNIAPIALNAODESSPTTMMNTNOELEKTKQLRID 480
 421 atIKAKENOLSIAPNNYPSKNIAPIALNAODESSPTTMMNTNOELEKTKQLRID 480
 481 TDQVGNATYFENGRRVDFGSMKSEVLPOIOETTAIINGDOLNVEERIAVANS 540
 481 tdQVGNATYFENGRRVDFGSMKSEVLPOIOETTAIINGDOLNVEERIAVANS 540
 481 tdQVGNATYFENGRRVDFGSMKSEVLPOIOETTAIINGDOLNVEERIAVANS 540
 481 tdQVGNATYFENGRRVDFGSMKSEVLPOIOETTAIINGDOLNVEERIAVANS 540
 541 DPLETTKPDMLTEALKAIFGNEPENGLOYGKDIETFEFNFDOOTSINIKQALAEINA 600
 541 dPLETTKPDMLTEALKAIFGNEPENGLOYGKDIETFEFNFDOOTSINIKQALAEINA 600
 541 dPLETTKPDMLTEALKAIFGNEPENGLOYGKDIETFEFNFDOOTSINIKQALAEINA 600
 541 dPLETTKPDMLTEALKAIFGNEPENGLOYGKDIETFEFNFDOOTSINIKQALAEINA 600
 601 TNYIVVLDKIKILAKNMIIRDKRPHYDRNNIAGADESVKFAHREVINSSTEGILLNI 660
 601 tnyIVVLDKIKILAKNMIIRDKRPHYDRNNIAGADESVKFAHREVINSSTEGILLNI 660
 601 tnyIVVLDKIKILAKNMIIRDKRPHYDRNNIAGADESVKFAHREVINSSTEGILLNI 660
 601 tnyIVVLDKIKILAKNMIIRDKRPHYDRNNIAGADESVKFAHREVINSSTEGILLNI 660
 661 DKDIRKIISGYIYEIETPEGLKEVINDRDMNISLRODGTFTDFKYNKPLIYISN 720
 661 dkDIRKIISGYIYEIETPEGLKEVINDRDMNISLRODGTFTDFKYNKPLIYISN 720
 661 dkDIRKIISGYIYEIETPEGLKEVINDRDMNISLRODGTFTDFKYNKPLIYISN 720
 661 dkDIRKIISGYIYEIETPEGLKEVINDRDMNISLRODGTFTDFKYNKPLIYISN 720
 721 PMYKVVAVYAVTKENT 735
 721 pmyKVVAVYAVTKENT 735
 721 pmyKVVAVYAVTKENT 735
 721 pmyKVVAVYAVTKENT 735

RESULT 2

AAB47306
 AAB47306 standard; Protein; 764 AA.

AAB47306;

29-AUG-2001 (first entry)

Wild type B. anthracis protective antigen.

Lethal factor; LF; immunogen; LF4; protective antigen; PA; DNA vaccine; humoral; cell-mediated; immune memory response.

Bacillus anthracis.

Key	Location/Qualifiers
Peptide	1..29
Protein	/label= Signal peptide
Peptide	/note= "Not given in the specification"
Peptide	30..764
Peptide	/label= PA
Peptide	204..764
Peptide	/label= PCPA

WO200145639-A2.

28-JUN-2001.

21-DEC-2000; 2000WO-US34912.

22-DEC-1999; 99US-0171459.

(OHIS) UNIV OHIO STATE RES FOUND.
 (GALLI/) GALLOWAY D R.
 (MATE/) MATECZUN A J.

Galloway DR, Mateczun AJ;

WPI; 2001-408540/43.

N-PSDB; AAC86016.

Protecting animal against lethal infection with Bacillus anthracis, by administering wild type or mutated form of Bacillus anthracis lethal factor protein or its fragment or a nucleic acid encoding the mutated protein -

Claim 5; Fig 2; 33pp; English.

This sequence shows the B. anthracis protective antigen (PA). An immunogenic fragment of PA, PCPA, can be used to produce an immune response which protects an animal against lethal infection with Bacillus anthracis. DNA encoding the B. anthracis PA can be used in conjunction with DNA encoding the lethal factor (LF) in a DNA vaccine. Using a DNA vaccine which encodes the mutated LF protein or fragment alone or in combination with a DNA encoding the PA protein or its fragment, both components (humoral and cell-mediated) of the immune system are stimulated, which results in longer term immune memory response. The combined use of a mutated LF and PA gene or their fragments results in a higher level of immune response, as judged by overall serum antibody titers for LF and PA antigens, than the use of either LF or PA genes in separate immunizations.

Sequence 764 AA:

Query Match 97.0%; Score 3652; DB 22; Length 764;

Best Local Similarity 97.6%; Pred. No. 5.9e-237;

Matches 717; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

1 MKRRVLPIMLSTIVSSGNLEVIQAEVQENRLNESSSSOGILGYFSDLNFOA 60
 1 mkkrrvlpimlslstivssgnleviqaevgqenrlnessessgllgyfstdlnfqa 60

```
QY 61 PAVVTSSTGDLSPSSSELENIPSENOYFOSAIMSGFIKKKSDEYTFATSADNHVTMV 120
XX |||||||
PI 61 pmvtsstgdlspssseleenisengyfgsaiswgfikkkdeytfatsadnhvmv 120
Db 61 pmvtsstgdlspssseleenisengyfgsaiswgfikkkdeytfatsadnhvmv 120
QY 121 DDQEVINKASNNKIRLEKGRLYQIKIYORENPTKEGLDFKLYWDSQNKKEVISSDNL 180
XX |||||||
XX 121 ddqevinkasnnkirkirlekgrlyqikiygrempkteqgldfklywdsqnkkevi ssdnl 180
Db 121 ddqevinkasnnkirkirlekgrlyqikiygrempkteqgldfklywdsqnkkevi ssdnl 180
QY 181 QLEBELKOKSSNSRRKSTSGPTVPVPRDNDGIFDSLEVEGYTDVKNKRFLESPWISNIH 240
XX |||||||
XX 181 qlebelkoksnsrrkstrsgptvpvprdndgipdslevegylvdvknkrfle spw isnih 240
Db 181 qlebelkoksnsrrkstrsgptvpvprdndgipdslevegylvdvknkrfle spw isnih 240
QY 241 EKKGLATKYSPEKMWSTASPYSDPEKVTGRIDKNVSPRARHPVLAAYPIVHDMENIIL 300
XX |||||||
XX 241 ekkglatkyspekswstaspydsdpkvtgridknvspearhplvaaypivhdm eniil 300
Db 241 ekkglatkyspekswstaspydsdpkvtgridknvspearhplvaaypivhdm eniil 300
QY 301 SKNEDOSTQNTDSETRTISKNTSTSRHTSEVHGNAEVHANTSTSRHTSEVHGNAEVA 360
XX |||||||
XX 301 sknedostqntdsetrtiskntstsrhtsevhgnaevhantstsrhtsevhg naeva 360
Db 301 sknedostqntdsetrtiskntstsrhtsevhgnaevhantstsrhtsevhg naeva 360
QY 361 VALDHSLSLAGERTWAEGLNADTARLANANIRYVNTGAPLYNVLPPTSILVIGKQOTL 420
XX |||||||
XX 361 valdhslslageretwaelnadtarlananiryvntgaplynvlpptsilv igkqotl 420
Db 361 valdhslslageretwaelnadtarlananiryvntgaplynvlpptsilv igkqotl 420
QY 421 ARIKAKENOLSOILAPNNYPSKNLAPIALNADDFSSPTTMNNOELEKTKOLRLD 480
XX |||||||
XX 421 atikakenolsolilapnnypsknlapialnaddfsspttmnnoelektkol rld 480
Db 421 atikakenolsolilapnnypsknlapialnaddfsspttmnnoelektkol rld 480
QY 481 TDQVGNIAATYFENGVRVDTGSNMSEVLPOIETTARIIFNGKDLNVERRIAIVNPS 540
XX |||||||
XX 481 tdqvgniaatynfengvrvdsgsnmsevlpoiettariifngkdlnverria ivnps 540
Db 481 tdqvgniaatynfengvrvdsgsnmsevlpoiettariifngkdlnverria ivnps 540
QY 541 DPETTRKPMTLKEALKIAGFENEPNGNLOYGKDIETEPFNFDOOTSQIKNOELAE 600
XX |||||||
XX 541 dpettrkpmtlkealkiagfenepngnloygkdietepfnfdootsqiknoel ae 600
Db 541 dpettrkpmtlkealkiagfenepngnloygkdietepfnfdootsqiknoel ae 600
QY 601 TNYTYVLDRKIKLAKNMNILLRDKRFHYDRNNIIVGADSVYKAEHREVINSSTEGLLNT 660
XX |||||||
XX 601 tnytyvlrkiaknmnillrdrkrfhydrnniivgadesvykahrvinssste gl lnt 660
Db 601 tnytyvlrkiaknmnillrdrkrfhydrnniivgadesvykahrvinssste gl lnt 660
QY 661 DKDIRKILSGYIEIEDTEGLKEVINDRYDMLNISSLRQDKTFIDFKKYNDKLLPYISN 720
XX |||||||
XX 661 dkdirkilsgyieiedteglkevindr ydmlnisslrqdkgtfidfkkyndk lplyisn 720
Db 661 dkdirkilsgyieiedteglkevindr ydmlnisslrqdkgtfidfkkyndk lplyisn 720
QY 721 PNKVVNVYANTKENT 735
XX |||||||
XX 721 pnkvvnvyavtkent 735
Db 721 pnkvvnvyavtkent 735
RESULT 3
AA156960
ID AA156960 standard; protein; 763 AA.
XX
AC AA156960;
XX
DT 25-APR-2000 (first entry)
XX
DE B. anthracis TPA-PA protein.
XX
KW Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;
KW tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.
XX
OS Bacillus anthracis.
XX
PN WO200002522-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US15568.
XX
PR 10-JUL-1998; 98US-0092416.
XX
```

```
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
XX Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL.
XX
XX WPI; 2000-182165/16.
DR N-PSDB; AA256876.
XX
XX Recombinant DNA construct useful as vaccines for anthrax, in producing
PT host cells for analyzing the drugs and agents inhibiting anthrax -
PS disclosure; Page 32; 35pp; English.
XX
XX The invention provides a recombinant DNA construct that comprises a
CC vector and at least one nucleic acid (or its fragment) encoding a
CC combination of Bacillus anthracis proteins, selected from protective
CC antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
CC (PA with its secretory signals replaced with those of tissue plasminogen
CC activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
CC for anthrax and in producing infectious alpha virus particles. These
CC particles, expressing the B. anthracis proteins are useful also as
CC vaccines for anthrax. Host cells transformed with the construct are
CC useful for analyzing the effectiveness of drugs and agents that inhibit
CC anthrax or B. anthracis proteins. The present sequence represents a
CC B. anthracis TPA-PA protein.
XX
XX
SQ Sequence 763 AA:
Query Match 94.3%; Score 3551.5; DB 21; Length 763;
Best Local Similarity 95.2%; Pred. No. 3.3e-230;
Matches 698; Conservative 7; Mismatches 25; Indels 3; Gaps 1;
QY 3 KKKVLIPLMALSTILVSTGNTLEYIOAEVKEQENLMESSSSGGLGYFSDNFGAPM 62
XX ||| : : : |||
Db 5 krglccvlllccgavfvaas---evlqaevxqenllinesessqllgyfisdntfqpm 61
QY 63 VVTSSTGDLSPSSSELENIPSENOYFOSAIMSGFIKKKSDEYTFATSADNHVTMVDD 122
XX |||||||
Db 62 vvtsstgdlspssseleenisengyfgsaiswgfikkkdeytfatsadnhv tmmvdd 122
QY 123 QEVINKASNNKIRLEKGRLYQIKIYORENPTKEGLDFKLYWDSQNKKEVISSDNLQ 182
XX |||||||
Db 122 qevinkasnnkirkirlekgrlyqikiygrempkteqgldfklywdsqnkkevi ssdnlq 181
QY 183 PELKQKSSNSKKRSTSGAPTVPVPRDNDGIPDSLEVEGYTDVKNKRFLESPWISNIH 242
XX |||||||
Db 182 pelkqksnsrkkstrsgaptvpvprdndgipdslevegylvdvknkrfle spw isnih 241
QY 243 KGLTKYKSSPEKMWSTASPYSDPEKVTGRIDKNVSPRARHPVLAAYPIVHDMENIILSK 302
XX |||||||
Db 242 kgltkykspekswstaspydsdpkvtgridknvspearhplvaaypivhdm eniilsk 301
QY 303 NEDOSTQNTDSETRTISKNTSTSRHTSEVHGNAEVHANTSTSRHTSEVHGNAEVA 362
XX |||||||
Db 302 nedgstqntdsetrtiskntstsrhtsevhgnaevhantstsrhtsevhg naeva 361
QY 363 IDHSLSLAGERTWAEGLNADTARLANANIRYVNTGAPLYNVLPPTSILVIGKQOTL 422
XX |||||||
Db 362 idhslslageretwaelnadtarlananiryvntgaplynvlpptsilv igkqotl 421
QY 423 IKAKENOLSOILAPNNYPSKNLAPIALNADDFSSPTTMNNOELEKTKOLRLD 482
XX |||||||
Db 422 ikakenolsolilapnnypsknlapialnaddfsspttmnnoelektkol rld 481
QY 483 TDQVGNIAATYFENGVRVDTGSNMSEVLPOIETTARIIFNGKDLNVERRIAIVNPSDP 542
XX |||||||
Db 482 tdqvgniaatynfengvrvdsgsnmsevlpoiettariifngkdlnverria ivnpsdp 541
QY 543 LETTRKPMTLKEALKIAGFENEPNGNLOYGKDIETEPFNFDOOTSQIKNOELAE 602
XX |||||||
Db 542 lettrkpmtlkealkiagfenepngnloygkdietepfnfdootsqiknoel ae 601
QY 603 IYTYVLDRKIKLAKNMNILLRDKRFHYDRNNIIVGADSVYKAEHREVINSSTEGLLNDK 662
```

Db 602 |ytvd|k|k|l|n|a|k|m|i|l|t|r|k|r|f|h|y|d|r|m|i|a|v|a|d|e|s|v|v|k|e|h|r|e|v|i|n|s|t|e|g|i|l|n|i|d|k 661
QY 663 |D|R|K|I|L|S|G|Y|I|E|D|T|E|G|L|K|E|V|I|N|D|R|Y|D|M|L|N|I|S|S|R|O|D|G|K|R|P|D|K|R|P|D|K|R|P|L|Y|I|S|N|P|N 722
Db 662 |d|r|k|l|s|g|y|i|e|i|e|d|t|e|g|l|k|e|v|i|n|d|r|y|d|m|i|n|i|s|i|r|g|d|g|k|t|i|d|k|k|y|n|d|i|p|l|y|i|s|n|p|n 721
QY 723 |Y|K|V|N|Y|A|V|T|K|E|N|T 735
Db 722 |y|k|v|n|y|a|v|t|k|e|n|t 734

RESULT 4

AAR60179 standard; protein; 735 AA.
ID AAR60179 standard; protein; 735 AA.

AAR60179:

03-APR-1995 (first entry)

Protective antigen of Bacillus anthracis.

Anthrax: Bacillus anthracis; fusion protein; protective antigen;

protective antigen; cell killing; targeting; targeting; pathogen;

intracellular; HIV; human immunodeficiency virus; toxin.

Bacillus anthracis.

MO9418332-A.

14-FEB-1994: 94WO-US01624.

12-FEB-1993: 93US-0021601.

25-JUN-1993: 93US-0082849.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Atora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;

WPI: 1994-279753/34.

N-PSDB: AAQ70180.

Nucleic acid encoding anthrax toxin fusion protein - useful for targeting toxin to specific cells, eg for killing tumour cells or HIV-infected cells

PS Disclosure; Page 81-83; 124pp; English.

CC The sequence encoding the protective antigen of Bacillus anthracis
CC may be used in the construction of a nucleic acid which encodes a
CC fusion protein comprising the anthrax protective antigen binding
CC domain of the native anthrax lethal factor and a sequence encoding an
CC activity inducing domain of a second protein. The fusion proteins
CC are useful for the specific killing of tumour cells or the killing of
CC cells infected with intracellular pathogens, especially HIV.

XX Sequence 735 AA;

Query Match 93.5%; Score 3519; DB 15; Length 735;

Best Local Similarity 97.5%; Pred. No. 4.8e-228;

Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 30 EVKQENRLINSESSSGGLGYFSDLNFOAPMVVTSSTGDIPTSPSELENIPSENYF 89
Db 1 evkqenrlinseesssgglgyfsdlnfqpvmvtsstgdlptspselehipsennyf 60
QY 90 QSAIWSGEFIKVKKSDYTFATSAADNHVTMWDDEVINKASNSNKRLEKGRLYQIKIY 149
Db 61 qsalwsgfikkksdeytfatsadnhvtmwddevlnkaansnkrlrekgrlyqikqiy 120

QY 150 |Q|E|N|P|E|K|G|L|D|F|K|L|Y|T|W|D|S|Q|N|K|E|V|I|S|D|N|Q|L|P|E|L|K|O|K|S|N|S|R|K|R|S|T|S|A|G|P|T|V|P|D|R|D|N 209
Db 121 |g|r|e|n|t|e|g|i|d|f|k|l|y|w|i|d|s|q|n|k|e|v|i|s|d|n|q|l|p|e|l|k|k|s|n|s|r|k|r|s|t|s|a|g|p|t|v|p|d|d|n 180
QY 210 |D|G|I|P|D|S|L|E|V|E|G|T|V|V|V|K|K|R|F|L|S|P|W|I|S|N|I|E|K|K|L|T|Y|K|S|S|P|E|K|W|S|A|S|D|P|Y|S|D|F|E|V|T 269
Db 181 |d|g|i|p|d|s|l|e|v|e|g|t|v|v|v|k|k|r|f|l|s|p|i|s|n|i|e|k|k|l|t|y|k|s|s|p|e|k|w|s|a|s|d|p|y|s|d|f|e|v|t 240
QY 270 |G|R|I|D|K|N|S|P|E|A|R|H|P|L|V|A|A|Y|P|I|V|H|D|M|E|N|I|I|S|K|N|D|O|S|T|O|N|T|D|S|E|P|T|I|S|K|N|T|S|P|R|H|T 329
Db 241 |g|r|i|d|k|n|s|p|e|a|r|h|p|l|v|a|a|y|p|i|v|h|d|m|e|i|l|i|s|k|n|d|e|s|t|q|n|t|d|e|c|r|i|s|k|r|s|t|h|c 300
QY 330 |S|E|V|H|G|N|A|E|V|H|A|N|T|S|R|T|H|T|S|E|V|H|G|N|A|E|V|H|A|V|I|D|H|S|L|A|G|E|R|T|A|E|T|M|G|I|N|T|A|R|L 389
Db 301 |s|e|v|h|g|n|a|e|v|h|a|n|t|s|r|t|h|t|s|e|v|h|g|n|a|e|v|h|a|v|i|d|h|s|l|a|g|e|r|t|a|e|t|m|g|i|n|t|a|r|l 360
QY 390 |N|A|N|I|Y|V|N|T|G|T|A|P|R|Y|N|V|P|T|S|V|G|K|K|O|T|A|T|R|A|K|E|N|O|S|Q|I|A|P|N|N|Y|P|S|K|L|A|R|A 449
Db 361 |n|a|i|r|y|v|n|t|g|t|a|p|r|y|n|v|p|t|s|v|g|k|k|o|t|a|t|r|a|k|e|n|o|s|q|i|a|p|n|n|y|p|s|k|l|a|r|a 420
QY 450 |I|N|A|Q|D|F|S|T|P|T|N|Y|N|O|F|L|E|K|R|Q|L|R|D|D|Q|Y|G|N|A|T|Y|N|F|E|N|G|H|V|R|V|D|G|S|M|S|E|V 509
Db 421 |i|n|a|q|d|f|s|t|p|t|n|y|n|o|f|l|e|k|r|q|l|r|d|d|q|y|g|n|a|t|y|n|f|e|n|g|h|v|r|v|d|g|s|m|s|e|v 480
QY 510 |L|P|Q|I|E|T|R|A|R|I|E|N|G|K|D|I|N|V|R|R|I|A|A|V|N|P|S|D|L|E|T|R|K|P|M|T|L|K|A|L|K|I|A|G|F|N|E|P|N|G|L 569
Db 481 |l|p|i|q|i|e|t|r|i|l|i|n|g|k|i|n|v|r|i|a|a|v|n|p|s|d|l|e|t|r|k|p|m|t|l|k|e|a|i|a|g|f|n|e|p|n|l 540
QY 570 |Q|Y|O|G|K|D|I|E|P|D|P|N|P|O|O|R|S|O|N|K|N|O|L|A|E|L|N|A|T|N|T|V|D|K|I|K|L|N|K|N|I|L|R|D|R|F|H|D|R 629
Db 541 |q|y|g|k|d|i|e|p|d|p|n|p|o|o|r|s|o|n|k|n|o|l|a|e|l|n|a|t|n|t|v|d|k|i|k|l|n|k|n|i|l|r|d|r|f|h|y|r 600
QY 630 |N|N|I|A|V|G|A|D|E|S|V|V|K|E|A|R|E|V|I|N|S|T|E|G|L|L|N|I|D|K|R|I|S|G|Y|I|E|D|T|E|G|L|K|E|V|I|N|D|R|Y 689
Db 601 |n|i|a|v|g|a|d|e|s|v|v|k|e|h|r|e|v|i|n|s|t|e|g|i|l|n|i|d|k|r|i|k|l|s|g|i|e|i|d|t|e|g|l|k|e|v|i|n|d|r|y 660
QY 690 |D|M|L|N|I|S|S|R|O|D|G|K|R|P|D|K|R|N|D|L|P|L|Y|I|S|N|P|N|K|V|N|Y|A|V|T|K|E|N|T 735
Db 661 |d|m|i|n|i|s|i|r|g|d|g|k|r|f|i|d|k|k|y|n|d|i|p|l|y|i|s|n|p|n|y|a|v|t|k|e|n|t 706

RESULT 5

AAY56959 standard; protein; 736 AA.

AAY56959:

25-APR-2000 (first entry)

B. anthracis MAT-PA protein.

CC Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;
CC tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.

XX Bacillus anthracis.

XX WO200002522-A2.

XX 20-JAN-2000.

XX 09-JUL-1999: 99WO-US15568.

XX 10-JUL-1998: 98US-0092416.

XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.

XX Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;

XX WPI: 2000-182165/16.

XX N-PSDB: AAZ56875.

XX Recombinant DNA construct useful as vaccines for anthrax, in producing
XX host cells for analyzing the drugs and agents inhibiting anthrax

XX Disclosure; Page 34; 35pp; English.

PS
XX
CC The invention provides a recombinant DNA construct that comprises a
CC vector and at least one nucleic acid (or its fragment) encoding a
CC combination of Bacillus anthracis proteins, selected from protective
CC antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
CC (PA with its secretory signals replaced with those of tissue plasminogen
CC activator) and PA63 (63 KDa PA). The construct is useful as a DNA vaccine
CC for anthrax and in producing infectious alpha virus particles. These
CC particles, expressing the B. anthracis proteins are useful also as
CC vaccines for anthrax. Host cells transformed with the construct are
CC useful for analyzing the effectiveness of drugs and agents that inhibit
CC anthrax or B. anthracis proteins. The present sequence represents a
CC B. anthracis MAT-PA protein.

XX
XX Sequence 736 AA;

Query Match 93.5%; Score 3519; DB 21; Length 736;
Best Local Similarity 97.5%; Pred. No. 4,8e-228;
Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

OY 30 EYKQENRLNSESSESSQGLGYFFSDLNFOAPMVVTSSTTGDISPSELENIPSENOYF 89
DB 2 evkqenrlinesesssgllgyffsdlnfqapmvvtssttgdispselenipsengyf 61
OY 90 QSAIWSGFIKVKKSDEYTFATSADNHVTMWVDOEVINKASNSNKIRLEKGRLYQIKIY 149
DB 62 qsalwsgfikvkksdeytfatsadnhvtmwvdoevinkasnsnkirlekgrlyqikly 121
OY 150 QRENPTFKGLDFKLYWTDSONKKEVISDNLQIPELKOKSSNRKRSTAGTPVPDRN 209
DB 122 qrenptekglidfklwtdsqnkkvissdnlqipelekqssnrkrstagsptvpdrn 181
OY 210 DGIPDSLEVEGYVDVKNKRTFLSPWISNIHEKKGLTKKSSPEKMTASDPYSDEKYT 269
DB 182 dgipdslevegylvdvknkrtflspwismhekkgltkksspekmtasdpysdefkyt 241
OY 270 GRIDKNVSPARHPPLVAAPYIVHVDMENILSKNEDOSTONTSEFTISKNTSTSTHT 329
DB 242 gridknvspearhplvaapyivhvdmenilsknedqstntseftiskntststht 301
OY 330 SEVHGAENVHANTSTSTHTSEVHGAENVHVAIDHSLIAGERTVAETMGLTADTARL 389
DB 302 sevhgaenvhasftststhtsevhgaenvhvaidsliagertvaetmgltdadtarl 361
OY 390 NANIRVTNTGTAPYIVNVLPFTSLVGNQTLATIKAKENOLSOILAPNNYPSKNLAPIA 449
DB 362 nanirvntgtaplyivnvlpftslvlgknqtlatikakengolsilapnnypsknlapia 421
OY 450 LNAQDFFSSTPTTMVNOFLELEKTRQRLTDQVYGNATYVFNENGRVAVDTSNMSEV 509
DB 422 lnaqdfssstpttmvnoflelektqrltdqvynatyvfnengrvavdtsnmsew 481
OY 510 LQIQIOTTRIIIFNGKDLNLYERRIAAVNSDPLETTKPMTLKEALKIAFGNERPGLN 569
DB 482 lpiqiottriiifngkdlnlyerriaavnsdplectkpmtlkealkiafgnepsn 541
OY 570 QYQKQDITFEFDFNQOTSONIKNOQLAELNATNYIVLQIKILAKAMNIIIRKRFHYDR 629
DB 542 qyqgkdtfefdfnqotsoniknoqlaelnatnyivldkiklakamniirkrfhydr 601
OY 630 NNIANGADESVKAEHREVINSSTEGLLNIDKIRKILSGYVEIETEGLEKVIENDRY 669
DB 602 nniavgadesvvkeahrevinsstegllnidkirkilsgyveieteglekeviendry 661
OY 690 DMNISLSRODGKTFIDFKKYNKLPYISNPNKYKNVAVTKENT 735
DB 662 dmnislsrtdgkftidfkknkldpyisnpyknvavtkent 707

RESULT 6

AAR60183
ID AAR60183 standard; Protein; 903 AA.

XX
AC AAR60183;
XX
DT 04-APR-1995 (first entry)
XX
DE PA(1-725)-----Human CD4 fusion protein coding sequence.

XX Anthrax; Bacillus anthracis; fusion protein; lethal factor;
KW protective antigen; cell killing; targeting; targeting; pathogen;
KW intracellular; HIV; human immunodeficiency virus; toxin;
KW pseudomonas; exotoxin.

XX
OS Bacillus anthracis.
OS Homo sapiens.

XX
PN W09418332-A.

XX
PD 18-AUG-1994.

XX
PF 14-FEB-1994; 94WO-US01624.

XX
PR 12-FEB-1993; 93US-0021601.

XX
PR 25-JUN-1993; 93US-0082849.
(USSH) US DEPT HEALTH & HUMAN SERVICES.

XX
PI Atora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;

XX
DR WPI; 1994-279753/34.

XX
DR N-PDSB; AAO70184.

XX
PT Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumor cells
or HIV-infected cells

XX
PS Disclosure; Page 100-103; 124pp; English.

XX
CC This sequence is a fusion protein comprising amino acid residues
CC 1-725 of the anthrax protective antigen protein and residues 1-178
CC of human CD4, the portion which binds to gp120 on HIV infected
CC cells. Such fusion proteins may be useful for the specific killing
CC of tumor cells or the killing of cells infected with intracellular
CC pathogens, especially HIV, depending on their components.

XX
SQ Sequence 903 AA;

Query Match 93.5%; Score 3519; DB 15; Length 903;
Best Local Similarity 97.5%; Pred. No. 6,5e-228;
Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

OY 30 EYKQENRLNSESSESSQGLGYFFSDLNFOAPMVVTSSTTGDISPSELENIPSENOYF 89
DB 1 evkqenrlinesesssgllgyffsdlnfqapmvvtssttgdispselenipsengyf 60
OY 90 QSAIWSGFIKVKKSDEYTFATSADNHVTMWVDOEVINKASNSNKIRLEKGRLYQIKIY 149
DB 61 qsalwsgfikvkksdeytfatsadnhvtmwvdoevinkasnsnkirlekgrlyqikly 120
OY 150 QRENPTFKGLDFKLYWTDSONKKEVISDNLQIPELKOKSSNRKRSTAGTPVPDRN 209
DB 121 qrenptekglidfklwtdsqnkkvissdnlqipelekqssnrkrstagsptvpdrn 180
OY 210 DGIPDSLEVEGYVDVKNKRTFLSPWISNIHEKKGLTKKSSPEKMTASDPYSDEKYT 269
DB 181 dgipdslevegylvdvknkrtflspwismhekkgltkksspekmtasdpysdefkyt 240
OY 270 GRIDKNVSPARHPPLVAAPYIVHVDMENILSKNEDOSTONTSEFTISKNTSTSTHT 329
DB 241 gridknvspearhplvaapyivhvdmenilsknedqstntseftiskntststht 300

Oy	330	SEVGNNAEVHANTSTSRKHHSEVNGNAEVAHVAADHSLSLAGERTAAETMGLTADTARL	389
Oy	330		
Db	301	sevgnaevhasffidlgsvsaagfsnssstvaldhs1slager1waeltmg1nltadtari	360
Oy	390	NANIEYVVTGAPLYNNVLPPTSLVLGKKNQTLATIKAKENQLSOILAPNNYPSKNIAPIA	449
Db	361	nanirvnttcgapiyrvnlptslvlgknqtlatlakenglsqilapnmypsknlapia	420
Oy	450	LNAODFSSPTITMNYNOFLELEKTKQLRLDPTDQVYGNIAFYNFENGRRVDTGSMWSEV	509
Db	421	lnagdfssptitlmnyqfielektkqlrltdqdvyniatytfengrvvdtgsmwse	480
Oy	510	LPOIETARIRIFNGKDLNLVERRIAAVNSDPLEETKKPMTLKEALKIFGNEENGNI	569
Db	481	lpqietetarilrfngkdlnlverriaavnsdpdlettkpmtlkealkiafgneengnl	540
Oy	570	OYOGKDIPEFPNFDQOSTQIRKNOQLAELNATNITYVLDKIKINAKMNIILRDKREHYOR	629
Db	541	qyqgkdiefefnfdqgstqirknqlelnatnityvlidikiklnakmnilrdkrrhydr	600
Oy	630	NNIANGADESVVKEAHREVINSSTEGLLNIDKDIRKILSGYIYEIEDPEGLKEVINDRY	689
Db	601	nniavgadesvvkeahrevinssteglllnidkdirkilsgylveledteglkevindr	660
Oy	650	DMLNSSLRDQGKTFIDFKKRNKDLPLIYSNPNTKVVAVVTKENT	735
Db	661	dmlnsslrdqgkttfidfkkyndklpiylisnpnykvnyavtkent	706
RESULT 7			
ID	AAR60193	standard; Protein: 719 AA.	
XX	AAR60193:		
XX	04-APR-1995	(first entry)	
DE	Modified protective antigen of Bacillus anthracis.		
XX	XX	Antitoxin; Bacillus anthracis; fusion protein; lethal factor;	
KW	XX	protective antigen; cell killing; targeting; targeting; pathogen;	
KW	XX	intracellular; HIV; human immunodeficiency virus; toxin.	
XX	OS	Bacillus anthracis.	
XX	PN	MO9418332-A.	
XX	PD	18-AUG-1994.	
XX	PF	14-FEB-1994; 94MO-USO1624.	
XX	PR	12-FEB-1993; 93US-0021601.	
XX	PR	25-JUN-1993; 93US-0082849.	
XX	PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
XX	PI	Arora N, Klimpel K, Leppia SH, Nichols PJ, Singh Y;	
XX	DR	WPI: 1994-279753/34.	
XX	DR	N-PSDB; MA070189.	
PT	Nucleic acid encoding anthrax toxin fusion protein - useful for		
PT	targeting toxin to specific cells, eg for killing tumour cells		
XX	or HIV-infected cells		
PS	Example 6; Page 114-115; 124pp; English.		
CC	The sequence encoding the protective antigen of Bacillus anthracis		
CC	may be used in the construction of a nucleic acid which encodes a		
CC	fusion protein comprising the anthrax protective antigen binding		
CC	domain of the native anthrax lethal factor and a sequence encoding an		
CC	activity inducing domain of a second protein. Such fusion proteins		
CC	are useful for the specific killing of tumour cells or the killing of		

ID	AAV56961 standard: protein: 569 AA.	AAV56961	Score 3342; DB 15; Length 719; Best Local Similarity 92.8%; Pred. No. 3,66-216; Matches 659; Conservative 6; Mismatches 216; Indels 24; Gaps 22;
CC	cells infected with intracellular pathogens, especially HIV,		
CC	depending on the second component. The protective antigen and other		
CC	toxins require proteolytic cleavage to acquire activity. Since some		
CC	cells infected with an intracellular pathogen possess an active		
CC	protease with quite a narrow substrate specificity e.g. HIV, the		
CC	protease cleavage site found in the native toxin is replaced with an		
CC	intracellular pathogen specific protease site (see AAR60184-88). The		
CC	protease in cells that are infected with an intracellular pathogen		
CC	cleaves the modified toxin which is then rendered active and kills		
CC	the cell. This sequence is a modified Bacillus anthracis protective		
CC	antigen which has the amino acids originally at positions 162-171		
CC	replaced with the HIV protease cleavable sequence described in		
CC	AAR60186.		
XX			
S0	Sequence 719 AA:		
QY	Query Match 88.8%; Score 3342; DB 15; Length 719;		
Db	Best Local Similarity 92.8%; Pred. No. 3,66-216;		
	Matches 659; Conservative 6; Mismatches 216; Indels 24; Gaps 22;		
QY	30 EYKQNRLLNESSSOGLLGYFSDLNQAAWVYTSSTGGLSPSSLENIPEENQF 89		
Db	1 EVKGEIRLLNESSSSGGLLYYISDILFQPMVYVTSSTGGLSPSSLENIPEENQF 60		
QY	90 QSAINSGEILVKKSDDEYFATFSADNHVTWVDDQEVINKASNMKIRLEKGLYQIKIY 149		
Db	61 GSAIVSGFIVKKSDEYFATFSADNHVTWVDDQEVINKASNMKIRLEKGLYQIKIY 120		
QY	150 QRENTPEKGLPKLWYTSQNKKEVYSSDNLQPELKQKSSNS---KKRRTSGAPYR 205		
Db	121 GRENTEKGLPKLWYTSQNKKEVYSSDNLQPELKQKSSNTATIMMGRGNFLLGPTVP 180		
QY	206 DEDNNGIPDSLEVEGYTVDVKKRPFELSPWISINIEKKGLTYKSSPEKMSAPSPYDP 265		
Db	181 dtdndgipdsleveytydvkkrpfelspwlsinekkgltyksspekmsaaspysdf 240		
QY	266 EKVYTRIDKNVSPEARHPLVAAVPIVHVDMEIILSKNEDOSTQNTDSEFTISKNTST 325		
Db	241 ekvtgridknvspearhplvaavpivhvdmeiilsknedogstqntdsetrlskntsts 300		
QY	326 RHTSEVHGNAVHNATSTRHTSEVHGNAEVHVAIDHSILAGERTWATGMLNTAD 385		
Db	301 rhtlsevhgnaevhnasftdldgsvsagfnsnssvtaidhsilageretwaelmgtad 360		
QY	386 TARLANIRYVVTGAPRIYVNPVPTSLVGLKMQTLATIKAKENQSLQILAPNNYPSKNL 445		
Db	361 tarlanirlyvvtgapiyvnvptslvlgkmtlatikakengsqilapnnypsknl 420		
QY	446 APIALNADDFSSPTITMNYNOFLELEKTQRLDPTDOYVGNIAIATVNFENGHVRVDTGSN 505		
Db	421 apialnaddfssptitlm-----ygnlatcyfenghvrvrdtgsn 460		
QY	506 WSEVLPQIOETTARIIFNGKDLNLVBRRAVNPSPDLTTPMDTLKALKIAGFNPR 565		
Db	461 wsevlpqioetatarilifngkdlnlvbrtaavnpdpdlttkpmtlkxalialafgnpr 520		
QY	566 NQNLQYOGDTEPDPNPDOOTSONIKNDLAEIANTVTVLDKIRLANKNMILLRDRF 625		
Db	521 ngnlqygqdtelfdnfntqgtsqnlknqlealnactlyvtldkrlaknmilltrdkf 580		
QY	626 HEDRNINAVGADSVYKKEAHREVINVSFTEGLLNTIDKIRLLSGYVIEEDTEGLKEVY 685		
Db	581 hdnrmnavigadesvykkehrevlnsstegllnldkdrkllsgyiveedteglkevi 640		
QY	686 NDRYMLNTISSLRDQKTFIDFKKYNDKLPLVYISNPNKRVNVYAVTKENT 735		
Db	641 ndrlymlntisslrqdgkftidfkkyndkrlplyisnpnyvnyavtckent 690		

XX AAY56961;
 AC
 XX
 DT 25-APR-2000 (first entry)
 DE B. anthracis PA63 protein.
 XX
 KM Bacillus anthracis protein: protective antigen; PA; MAT-PA; TPA-PA;
 KM tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.
 XX
 OS Bacillus anthracis.
 XX
 PN WO200002522-A2.
 XX
 PD 20-JAN-2000.
 XX
 PF 09-JUL-1999; 99WO-US15568.
 XX
 PR 10-JUL-1998; 98US-0092416.
 XX
 PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
 XX
 PI Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;
 XX
 DR WPI: 2000-182165/16.
 DR N-PSDB; AA556877.
 XX
 PT Recombinant DNA construct useful as vaccines for anthrax, in producing
 PT host cells for analyzing the drugs and agents inhibiting anthrax -
 XX
 PS Disclosure; Page 35; 35pp; English.
 XX
 CC The invention provides a recombinant DNA construct that comprises a
 CC vector and at least one nucleic acid (or its fragment) encoding a
 CC combination of Bacillus anthracis proteins, selected from protective
 CC antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
 CC (PA with its secretory signals replaced with those of tissue plasminogen
 CC activator) and PA63 (63 KDa PA). The construct is useful as a DNA vaccine
 CC for anthrax and in producing infectious alpha virus particles. These
 CC particles, expressing the B. anthracis proteins are useful also as
 CC vaccines for anthrax. Host cells transformed with the construct are
 CC useful for analyzing the effectiveness of drugs and agents that inhibit
 CC anthrax or B. anthracis proteins. The present sequence represents a
 CC B. anthracis PA63 protein.
 CC
 CC Sequence 569 AA;
 XX
 SQ

Query Match 70.8%; Score 2664; DB 21; Length 569;
 Best Local Similarity 96.7%; Pred. No. 8.8e-171;
 Matches 521; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 197 STSAGPTVPDRNDGIPDSLEVEGYTVVKNKRTFLSPWISNIHEKKGLTKYKSSPERKS 256
 DB 2 stsaagptvpdrndgipdslevegtyvdknkrtflspwlsnhkxgltkykaspekws 61
 QY 257 TADDPYSDPEKKVGRITDKNNSPFAHPLVAAYPIVHDMENIILSKNEDOSTOTWDESTR 316
 DB 62 tasdpydsdekkvgritdknnspfarhplvaaypivhdmennilsknedstgtqtdsetr 121
 QY 317 TIKKNSTSTRHTSEVHGNAEVAHANTSTSRHTSEVHGNAEVAHVALDHSLSLAGERTWA 376
 DB 122 tikknststrhtsevhgnaevahantstsrhtsevhgnaevahvaldhsldhsldagertwa 181
 QY 377 ETMGLNTADTARLANIRVYNTGTAPRYNVLPTTSLVLGKNQTLATIKAKENQSLIAP 436
 DB 182 etmglntadtarlanirvnyntgtaprynvlpstslvlgknqtlatlakengsqliap 241
 QY 437 NNTYPSKNTLAPIALNADODSSPTITMNYQOFLELETKQLRLDPTDOYGINATYFENG 496
 DB 242 nntypskntlapialnadodssptitmnynqoflelektqlrltdptdyginatynfeng 301
 QY 497 RVRVDFTGSNMSEVLPQIOETTARIIFNGKDLNLVERLIAAVNPSPDELTTRKPDWTLKEAL 556

DB 302 rrvvdftgsnmsevpqioettarilifngkdlnlverliaavnpdsetctpdmntlkeal 361
 QY 557 KIAFEPNEPNGNIQYCGKDITREFDFNFDOQTSQNIKNQLAELMANTVYTLDKIKLNAM 616
 DB 362 kiafepnepngnlqyqgkditefdnfdfqtsqnlknqaelmactnlytvidkiklnaem 421
 QY 617 NIILROKRFHNDNRNNIANGADESVYKKAHREVINSSTGGLNLNDKQIRKILSGIVAE 676
 DB 422 niilrdkrfhydrnlniavadesvvykkahevlnsteglllnldkdkrlkllsgylve 481
 QY 677 DTEGLKEVINDRYDMLNISSLRQDGKPTIDFKKYNDRKPLVYISPNKYVNTYAAVTKEMT 735
 DB 482 dteglkevindr ydmlnisslrqdgktidfkknkdrplvyspnkyvnyavtke 540

RESULT 9
 AAR91239
 ID AAR91239 standard; Protein; 884 AA.
 XX
 AC AAR91239;
 XX
 DT 14-AUG-1996 (first entry)
 XX
 DE B. cereus VIP1(a) insect-specific protein.
 XX
 KM Pesticide; insecticide; biological control agent; Lepidoptera;
 KM Coleoptera; transgenic plant; maize; insect resistance;
 KM western corn rootworm; Diabrotica virgifera virgifera; VIP.
 XX
 OS Bacillus cereus strain AB78 (NRRL B-21058).
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..33
 FT /label= sig_peptide
 XX
 PN W09610083-A1.
 XX
 PD 04-APR-1996.
 XX
 XX 27-SEP-1995; 95WO-EP03826.
 PF 05-JUN-1995; 95US-0463483.
 PR 28-SEP-1994; 94US-0314594.
 XX
 PA (CIBA) CIBA GEIGY AG.
 XX
 PI Carr B, Desai NM, Duck NB, Estruch JJ, Kostichka K;
 PI Koziel MG, Mullinma, Nye GJ, Warren GW;
 XX
 DR WPI: 1996-200921/20.
 DR N-PSDB; AAT13940.
 XX
 PT Bacillus strain producing insecticidal protein during vegetative
 PT growth - used in the control of Lepidoptera and Coleoptera pests
 XX
 PS Claim 15; Page 121-124; 242pp; English.
 XX
 CC Insect-specific protein VIP1(a) (AAR91239) of Bacillus cereus AB78
 CC shows activity against Diabrotica spp. pests such as the
 CC western corn rootworm. It is encoded by the VIP1(a) gene (AAT13940)
 CC isolated from a cosmid clone of AB78. VIP1(a) can be expressed in
 CC e.g. bacterial hosts to provide biological control agents having
 CC increased activity or target range, or can be expressed in transgenic
 CC plants, esp. maize, to improve insect resistance. It is preferably
 CC expressed as a fusion protein (see also AAR91245) with auxiliary protein
 CC VIP2(a) (AAR91238).
 CC
 CC Sequence 884 AA;
 XX
 SQ

Query Match 20.9%; Score 787.5; DB 17; Length 884;
 Best Local Similarity 26.7%; Pred. No. 1.9e-44;

Matches	248;	Conservative	134;	Mismatches	291;	Indels	255;	Gaps	32;
OY	1	MKKRKVLIPLMALSTILVSS--TGNLEVIOAEVK-----OENRLNSESSESSOGILGY	51						
DB	1	mkmmkkkiasvvtctllapmfingvnayadsktngistctgknq---qkemdrgkyllyg	57						
OY	52	YPSDLNFQAPMNVVTSSTGDLSPSSLELN--IPSENOYFQSAIMSGFIKVKKSDEYTPA	109						
DB	58	yfkgykdf-snlmfaptctdstllydqgtanklldkkgqgygsirwylglsksetgdfth	116						
OY	110	TSADNRHVTMWDQEVINKASNSNKRILEKGRLOYKIOYQENPEPEKLD-----FKL	163						
DB	117	lseedegallieingklismngkekgyvnhlekylvpikieygsd--tkfidskktfkell	174						
OY	164	YMTDSQNKKEVIVSSDNLQJPELKOKS-----NSRKKRSTAGPTVPDRND	210						
DB	175	fklidsgnqgqygqdelirpefnfkkesgeflakpsklnlftqkmkreided---ldtdgd	231						
OY	211	GIPDSLEVEGYTVQVKNKRTFLSPWISNHEKKGLTKYKSPSEKWSSTASDPYSDEKVTG	270						
DB	232	sipdliweengytl-----qnrliavkwddsl-aaskgytlkfvsnpleshtvgdpytdyekaar	286						
OY	271	RIDKNVSPARHPLVAAYIVHVDMENIILSKNEDOSTONTSERTISKNTSTSTHTS	330						
DB	287	dlidlnaketfnplvaafpsvvnsmekvllspnenls-----	323						
OY	331	EVHGNAEVHANTSTSRHTSEVHGNAEVH-----AVAIHDSLSLAGEPTAETMG--	380						
DB	324	---nsveshstcswsyntlegasveaglgpkglstfgsvnyqhselvaege--wgstsgnt	378						
OY	381	--LNTADTARLANIRYVNTGTAPITVNIPTSLVIGKQOTLATIKAKNOISQILAPNN	438						
DB	379	sqgfinaesagylaanvrynmvgcgaigdyvkrptstfvl-mndlatlataknsfahlsipge	437						
OY	439	YPSKNIAPIALNADDFESTPTIMVYNQFLELEKTKOLRITDOYGMATYINPENGAV	498						
DB	438	syppkgygngjaltsmdfnshpnltnkkkqvahllnnkpmmltetngtdg---vykikdtng	494						
OY	499	RVDTSNMSSEVLPOIETARTIIFNGKDLNLVERIRIAAVNPSDPLETTTPTDMLKEALKI	558						
DB	495	nltvgewngvliqglkakctasllvdgge--rvaekrvaakdyenpedkt-psltlkdaalkl	552						
OY	559	AF--GFNENGNLQYQGRKITEFD--NPDQOTSONIKRQNLDEL-----NATNIITYLDK	609						
DB	553	syppdelkelegillykxkpklyessvmtlyldentakvevtqjndtgcikfkdvshlydv---	609						
OY	610	IKLNKAKNMLIRDKRPFHYDRNNIIVAGADESVYKFEAHRVINSSTEG-----LL	657						
DB	610	-kltpkmvntlk-lsillydn-----aesndnsigkvtchntnlvsgngngkkqysnnpdantl	664						
OY	658	LN-----	659						
DB	665	lntdagelknrdyaislysmkseknitgceltdgeilyptlcttktvnvnkdykridliah	724						
OY	660	-----IDKDIKRLISGIYIEIETEGCL	681						
DB	725	nksnpisnlhktndetlflfwdldisltcvasikpenltdselkqlyrsrygikled--g	782						
OY	682	-----KEVIND-----RY-----DMLNTISSLRQDCKTPI	705						
DB	783	ltdkkgqglhygefineasfnlsieplqnyvtykyevtysselsignvstclsesdklykdgltkf	842						
OY	706	DEKKY--NDKLPILYISNPNYKVVYAVT	731						
DB	843	dftkyskneqglfydsglnwdfnlnait	870						

DT	15-APR-1998	(first entry)
XX		
DE	B. cereus VIP1A(a) protein sequence.	
XX		
KW	Vegetative insecticidal protein; Bacillus cereus strain AB78; plant;	
KW	insect; Sesamia nonagrioides; maize; corn borer; toxin.	
XX		
OS	Bacillus cereus.	
XX		
PN	W09726339-A1.	
XX		
PD	24-JUL-1997.	
XX		
PF	23-DEC-1996; 96WO-EP05828.	
XX		
PR	15-JAN-1996; 96GB-0000786.	
XX		
PA	(NOVS) NOVARTIS AG.	
XX		
PI	Gay PB;	
XX		
DR	WPI: 1997-385342/35.	
XX		
DR	N-PSDB; AAT733994.	
XX		
PT	Protecting plants against insects of the genus Sesamia using	
PT	Bacillus toxic proteins - applied directly or expressed as	
PT	heterologous protein by the plant, also transgenic plants expressing	
XX	both Cry and VIP type toxins	
PS	Claim 8; Page 35-39; 168pp; English.	
XX		
CC	This is the amino acid sequence of the 100 kD vegetative insecticidal	
CC	protein (VIP) 1A(a) from Bacillus cereus strain AB78. The protein can	
CC	be used in a new method for protecting plants, and their progeny, against	
CC	insects of the genus Sesamia by direct or indirect application to the	
CC	plant (or seed or growing area). The protein is especially useful to	
CC	protect maize plants against the Mediterranean corn borer	
CC	(S. nonagrioides).	
XX		
SQ	Sequence 884 AA;	
Query Match	20.8%; Score 783.5; DB 18; Length 884;	
Best Local Similarity	29.6%; Pred. No. 3.5e-44;	
Matches	234; Conservative 133; Mismatches 293; Indels 131; Gaps 29;	
OY	1	MKKRKVLIPLMALSTILVSS--TGNLEVIOAEVK-----OENRLNSESSESSOGILGY
DB	1	mkmmkkkiasvvtctllapmfingvnayadsktngistctgknq---qkemdrgkyllyg
OY	52	YPSDLNFQAPMNVVTSSTGDLSPSSLELN--IPSENOYFQSAIMSGFIKVKKSDEYTPA
DB	58	yfkgykdf-snlmfaptctdstllydqgtanklldkkgqgygsirwylglsksetgdfth
OY	110	TSADNRHVTMWDQEVINKASNSNKRILEKGRLOYKIOYQENPEPEKLD-----FKL
DB	117	lseedegallieingklismngkekgyvnhlekylvpikieygsd--tkfidskktfkell
OY	164	YMTDSQNKKEVIVSSDNLQJPELKOKS-----NSRKKRSTAGPTVPDRND
DB	175	fklidsgnqgqygqdelirpefnfkkesgeflakpsklnlftqkmkreided---ldtdgd
OY	211	GIPDSLEVEGYTVQVKNKRTFLSPWISNHEKKGLTKYKSPSEKWSSTASDPYSDEKVTG
DB	232	sipdliweengytl-----qnrliavkwddsl-aaskgytlkfvsnpleshtvgdpytdyekaar
OY	271	RIDKNVSPARHPLVAAYIVHVDMENIILSKNEDOSTONTSERTISKNTSTSTHTS
DB	287	dlidlnaketfnplvaafpsvvnsmekvllspnenls-----
OY	331	EVHGNAEVHANTSTSRHTSEVHGNAEVH-----AVAIHDSLSLAGEPTAETMG--
DB	324	---nsveshstcswsyntlegasveaglgpkglstfgsvnyqhselvaege--wgstsgnt

RESULT 12
 AAR91245
 ID AAR91245 standard: Protein, 1346 AA.
 AC AAR91245;
 XX
 XX 14-AUG-1996 (first entry)
 DT
 XX
 DE VIP2A(a) and VIP1A(a) fusion protein.
 XX
 KW Pesticide; insecticide; biological control agent; lepidoptera;
 KW Coleoptera; transgenic plant; maize; insect resistance;
 KW Western corn rootworm; Diabrotica virgifera virgifera; VIP.
 OS
 XX Bacillus cereus strain AB89 (NRRL B-21058).
 PN
 XX WO9610083-A1.
 PD
 XX 04-APR-1996.
 XX
 XX 27-SEP-1995; 95WO-EP03826.
 PF
 XX 05-JUN-1995; 95US-0463483.
 PR 28-SEP-1994; 94US-0314594.
 XX
 XX (CIBA) CIBA GEIGY AG.
 PA
 XX Carr B, Desai NM, Duck NB, Estruch JJ, Kostichka K;
 PI Kozel MG, Mullinsma, Nye GJ, Warren GW;
 XX
 DR WPI: 1996-200921/20.
 N-PSDB; AAT13944.
 XX
 XX Bacillus strain producing insecticidal protein during vegetative
 PT growth - used in the control of lepidoptera and Coleoptera pests
 XX
 PS Claim 30; Page 156-61.; 242pp; English.
 XX
 CC A fusion protein (AAR91245) is composed of the VIP2A(a) auxillary
 CC protein (AAR91238) and insect-specific VIP1A(a) protein (AAR91239) of
 CC Bacillus cereus strain AB89. It is the product of a fusion gene
 CC (AAT13944) constructed from the VIP coding sequences. The fusion
 CC protein can be expressed e.g. in bacterial cells for use as
 CC biological control agents having improved activity or target range,
 CC or in transgenic plants, esp. maize, to increase insect resistance.
 CC VIP1A(a) shows activity against western corn rootworm.
 CC
 XX
 SO Sequence 1346 AA;

Query Match 20.8%; Score 783.5; DB 17; Length 1346;
 Best Local Similarity 29.6%; Pred. No. 6.5e-44;
 Matches 234; Conservative 133; Mismatches 293; Indels 131; Gaps 29;

QY 1 MKRRKVLPLMALSTIIVSS--TGNLEVIQAQEVK-----QENRLNSESSESSOGLLGY 51
 DB 463 mkmkmlasavtctllapmfingvnayadsktngstctgkng---qkmdrkglylgy 519
 QY 52 YFSDLNFPAPMNVVSTTGDSLSPSELEN--IPSENOYFOASINSGFKYAKSDEYTPA 109
 DB 520 yfkgkdf-snlmftaptdstllydqfanklldkkqegyslrylglisqsetgdfth 578
 QY 110 TSADNHRVMTWDDDEVINKASNSNKIRLEKGRUYOIKIOYOENPTEKGLD-----PKL 163
 DB 579 lsedegatleingklisnkykgyvnhlekylipikleygsd--tkfnidcktkelkl 636
 QY 164 YWTDSONKKEVISSDNILQPLKOKRS-----NSRKRSTSAQPTVDRDRND 210
 DB 637 fkidsgnqpgqvqdeltnpfnkkesqelakpsklnlfctkmrledded---ldtdgd 693
 QY 211 GIPDSLEVEGTVVYKMKRTFLSPWISINHEKKGLTKYKSSPEKMSSTASDPYSDEKVTG 270

DB 694 sipdlweengylt-----gnrlavkwddsl-askgyckltvsnplshetvgydpyldyekaar 748
 QY 271 RIDKNVSPPEARHPVLAAYPIVHVDENILILSKNEDOSTGNTSETRTISKNTSTRTHTS 330
 DB 749 dldlnaketfnplvaafsvvsmekvllspenls----- 785
 QY 331 EYHGNAEYHANTSTSRKHTSEVHGNAEVH-----AVADHSLSLACERTWAETMG-- 380
 DB 786 ---nveshstswstwnltnegasveaglpkglsfgyvsnghsetvage--wgstgnt. 840
 QY 381 --LNTADTARLANIRYVNTGTAPITYNVLEPTSLVGLKQOTLATIKAKENOLSQILAPNN 438
 DB 841 sqfnlasagylaanvrynmgvgalvdvkrptsfyl-mdtlatlaksnsctalispge 899
 QY 439 YPSPKNLAPILNADDDFSEPTITMNYNOFLFELEKTKOLRLTDQOYGNATYNGENGV 498
 DB 900 syppkkgngiaitlsmddfshpiltlnkkqvdhllnnkpmlietngtdg---vykkikdng 956
 QY 499 RVDTSNMSSEVLPOIQETTARIIFNGKDLNLVERIAAVNPSDPLETTKPDMLKEALKI 558
 DB 957 nlvtgwegwvlgqlkaktasllvddge-rvaekrvaaakdyenpedkt-psltlkdaikl 1014
 QY 559 AF--GFNEPENGMLYOQGDITEFDF--NPDQOTSONIKQLAEL-----NATNITYVLDK 609
 DB 1015 sypdakelelegillyknpiyessvmtlyldentakvctkqldctgkfkdvshlydv--- 1071
 QY 610 IKLNKAKMILIDRKRFPHYDRNNIYAGADESVYKEAHREYINSSTEG-----LL 657
 DB 1072 -kltpkmvnlk-lsllydn---aesndnsigkwtcnlnlvsgngngkkyssnpanlct 1126
 QY 658 LNIID-----KDRIKILSGYIEIDTEGLKEVINDRYDMLNIS-SLRDGGTFIDFKKY 710
 DB 1127 lntdageklnkndrylislmyksekntgcetlidgeilyptlctkvtvnkndnykkrldiah 1186
 QY 711 NDKLPLYSNP 721
 DB 1187 mlk-----snp 1192

RESULT 13
 AAW19513
 ID AAW19513 standard: Protein, 1346 AA.
 AC AAW19513;
 XX
 XX 15-APR-1998 (first entry)
 DT
 XX
 DE B. cereus VIP1A(a)/VIP2A(a) fusion protein sequence.
 XX
 KW Vegetative insecticidal protein; Bacillus cereus strain AB78; plant;
 KW insect; Sesamia nonagrioides; maize; corn borer; toxin; fusion protein.
 XX
 XX Bacillus cereus.
 OS
 XX WO9726339-A1.
 PN
 XX 24-JUL-1997.
 PD
 XX 23-DEC-1996; 96WO-EP05828.
 PR 15-JAN-1996; 96GB-0000786.
 XX
 XX (NOVS) NOVARTIS AG.
 PA
 XX Gay PB;
 PI
 XX WPI: 1997-385342/35.
 DR N-PSDB; AAT73999.
 XX
 XX Protecting plants against insects of the genus Sesamia using
 PT Bacillus toxic proteins - applied directly or expressed as
 PT heterologous protein by the plant, also transgenic plants expressing

PT both Cry and VIP type toxins

XX Claim 8; Page 69-75; 168pp; English.

CC This is the amino acid sequence of a fusion protein comprising the
CC vegetative insecticidal proteins (VIP) 1A(a) and VIP2A(b) from *Bacillus*
CC cereus strain AB78. The fusion protein can be used in a new method for
CC protecting plants, and their progeny, against insects of the genus
CC *Sesamia* by direct or indirect application to the plant (or seed or
CC growing area). The protein is especially useful to protect maize plants
CC against the Mediterranean corn borer (*S. nonagroides*).

XX Sequence 1346 AA;

Query Match 20.8%; Score 783.5; DB 18; Length 1346;
Best Local Similarity 29.6%; Pred. No. 6.5e-44;
Matches 234; Conservative 133; Mismatches 293; Indels 131; Gaps 29;

```

QY 1 MKRRKVLIPMALSTIIYSS--TGNLEVIOAEVK-----QENRLNSESSSOGILGY 51
DB 463 mkmkxkkliaavtctllapmflngnvaavadsktngistqkng---qkemdtkgylly 519
QY 52 YFSDLNFOAPMVVYSTTGDLSPSSLELN--IPSENOYFOSAIMSGFIKVKSDYEYFA 109
DB 520 yfkqkdf-snlmtfaptrdcltydgqtanklklkkgqgysitwlgigsketgdtfn 578
QY 110 TSAADNHTVMVNDQEVINKASNSKIRLEKGRLOYKIOYORENPTEKGLD-----FKL 163
DB 579 lseegallielnglllenkqekqkvnhlekylvplkleygsd--tkfnidsktfkkelkl 636
QY 164 YWTDSONKKEVSSDNQLPELKOKS-----NSRKKRSTSAQPTVPDRND 210
DB 637 fklidsqgqgqvqgdcltnpdefnfkqesqetlaxpsklnlftqkmkreided--ldtgdg 693
QY 211 GIPDSLEGEYTVDKMKRFLSPWISINIEKKGLTKYKSSPEKMTASDYSPEKKTG 270
DB 694 slpdlwengytl-----gnrlavkwdsj-askytklftvnplshvqgdtycltckear 748
QY 271 RIDKNVSPEARHPLVAAYPIYHVDMEIILSKNEDOSTQNTDSEPTISKNTSTSRHTS 330
DB 749 dlqlsnaeketfnplvaafpsvnmekvllspnenls----- 785
QY 331 EVHGNAEVHNTSRTSRHTSEVHGNAEVH-----AVATDHSLSLAGEETMAETMC-- 380
DB 786 ---nsveshstwsyntegasyeaglgpkglisfgvsvngqhselvaqe--wglstqnt 840
QY 381 --LNTADTARLANIRYVNTGTAPIYVNLPTTSLVLGKNQTLATIKKENQLSLAPNN 438
DB 841 sgntaaagylnanvrynmvgtaigvdkptstfvl-ndtialitakstalnispge 899
QY 439 YFSKKNLAPLANAQDFFSTPTTMANYNQLEKTKQLRLTDQVYGNATYVENGSRV 498
DB 900 slpdkgnggjaltsmdfdnshphtltnkkydnllnnkpmmltqtdg---vyikdthg 956
QY 499 RYDTGSWMSRVLPQIOETTRIIIFNGKDLNVERRIAAVNSDPLETTKPMITKEALKI 558
DB 957 nlvtgagewngvlgikaktsasliwdgge-ryaekryvaakdyenpedkt-psltlklkdlk 1014
QY 559 AF--GFNEPUNGNLQYOGCKDTEFDF--NPDQOTSONIKNOIAEL-----NATNITYVL 609
DB 1015 sydelkeiegllyllykknpiyessvmtlyidentakevtkgndltctgfkdvshlydv--- 1071
QY 610 IKLANAKNMLIRDKRFHYDRNNINAVGADSEVYKAEHREVINSSTEG-----LL 657
DB 1072 -kltpkmvclt-lsilygd---aesndnsigkvtlnlvisgngnkqysnnpdantl 1126
QY 658 LNTD-----KDIRKIISGIVEIEDTEGLKEVINDRYDMINIS--SLRQDCKTIDRKY 710
DB 1127 lntdageklknrdyislzymkseknctgcetltdgeiylptlctkvnnkdmknykrlidiah 1186
QY 711 NDKLPLYSNP 721
DB 111

```

DB 1187 ntk-----snp 1192

RESULT 14

AAW46723 standard; Protein; 1346 AA.

AAW46723;

02-JUN-1998 (first entry)

VIP1A(a)/VIP2A(a) fusion protein.

Vegetative insecticidal protein; VIP; expression; maize; protection;
plant; *Ostrinia furnacalis*; Asian Corn Borer; Cry toxin; VIP toxin;
recombinant; *Bacillus thuringiensis*; transgenic plant; resistance;
insect attack; *Sesamia*; maize; cereal crop.

Synthetic.

Bacillus sp.

W09746105-A1.

11-DEC-1997.

27-MAY-1997; 97WO-EP02737.

06-JUN-1996; 96GB-0011777.

(NOVS) NOVARTIS AG.

Hunter B, Suwantaradon K, Dildewilligen WPM;

WPI; 1998-041787/04.

N-PSDB; AAV16172.

Claim 8; Pages 78-83; 175pp; English.

The present sequence represents the fusion protein of vegetative
insecticidal protein 1A(a) (VIP1A(a)) and VIP2A(a). The protein is
used in a method for protecting plants and their progeny against
damage caused by *Ostrinia furnacalis* (Asian Corn Borer). The
protein is directly or indirectly applied to the plant, plant seed or
growing area of the plant. Cry toxins can also be used in the same way,
in place of VIP toxins. The Cry or VIP toxins and genes are used,
especially inside recombinant B. cereus or B. thuringiensis strains, to
produce plants protected against Asian Borer pests. Transgenic plants
protected against Asian Corn Borer can be used to produce seed and
progeny also resistant to insect attack. Plants expressing both a
Cry-type and a VIP toxin gene can also protect against *Sesamia* pests.
The method and compositions are especially used for protecting maize
but may also be used to protect other cereal crops against Asian Corn
Borer attack.

Sequence 1346 AA;

Query Match 20.8%; Score 783.5; DB 19; Length 1346;
Best Local Similarity 29.6%; Pred. No. 6.5e-44;
Matches 234; Conservative 133; Mismatches 293; Indels 131; Gaps 29;

```

QY 1 MKRRKVLIPMALSTIIYSS--TGNLEVIOAEVK-----QENRLNSESSSOGILGY 51
DB 463 mkmkxkkliaavtctllapmflngnvaavadsktngistqkng---qkemdtkgylly 519
QY 52 YFSDLNFOAPMVVYSTTGDLSPSSLELN--IPSENOYFOSAIMSGFIKVKSDYEYFA 109
DB 520 yfkqkdf-snlmtfaptrdcltydgqtanklklkkgqgysitwlgigsketgdtfn 578

```

```
OY 110 TSADNHVTMMWDQEVINKASNSNKRIRLEKGRLYOIKIOYOBENPTREKLD-----FKL 163
DB 579 lsedegallieingklisnqkqkqvnhlekqkvlplkleygsd--tkfidskctkelk 636
OY 164 WYTDSONKKEVISDNLOJPELKOKS-----NSRKRKSTAGPTVPDRDND 210
DB 637 fklidqngqgvgqgdellrpefnkkesqeflakpsklnlftqkxmrelded---tdtdgd 693
OY 211 GIPDSLEVEGYTVDVKNKRTPLSPWISNIEKKGLTKYKSSPEKMSSTASDPYSDPEKVTG 270
DB 694 slpdlweengytl-----qnrliavkwddsl-askytklfvsnplshstlvgydpytdyekaar 748
OY 271 RIDKANSPEARHPVAAAYPIVHVDMENILTSKNEDOSTONTSERTISKNTSTSTHTS 330
DB 749 dldlsnaketlnplvaafpsvvnsmekvllspnenls----- 785
OY 331 EVHGNAEVHANTSTSRTHSEVHGNAEVH-----AAVHSHSLAGEERTWAETMG-- 380
DB 786 ---nveshststnwsytnlegasveaglsfgsvnyghsetvage--wgstsgnt 840
OY 381 --LNTADTARLANIRVYNTGTAPLYNVLPPTSLVLGKNQTLATIKAKENQSLIAPNN 438
DB 841 seqftasaqylaanvrynmgvgatdydvkptsfvlnndlatlftakensstalnspge 899
OY 439 YPSPNNLAPIALNADDESTPTIMYNOFLELEKTKOLRDTQDYGNIAIYFNENGRV 498
DB 900 sypkkgngqgialtsmddinspnlckkqvndllnkpmlleatngtdg---vykklkdtg 956
OY 499 RVDTGSNNSEVLPOIETARTIIFNGKDLNLERRIAAVNPSPDEPTTYPDMTLKEALKI 558
DB 957 nltvgewmgvqglkakkaasllvdgge-rraekrraadvapenpekt-psltlkdaalkl 1014
OY 559 AF--GFNENGULQYQKDIETEDF--NPDQOTSONIKQOLAEI-----NATNIYTVLDK 609
DB 1015 sypdelkelegllyknpkilyessvmtlyidentakevtqldntgkfkdvshlydv--- 1071
OY 610 ICLNKMNLIDKRPYHVRNNIAGADESVYKKAHREVINSTEB-----LL 657
DB 1072 -kltpkmvnlk-lsllydn---aesndnsigkwtlnltsqngnkkyssnnpdaant 1126
OY 658 INID-----KDIRKILSGYIIEDETEGLKEVINDRYDMLNIS-SLRDQKTFIDFKKY 710
DB 1127 lntdageklnkrdyylslsymkseknltgceltdgeilpylttktyvnvknkndykrldish 1186
OY 711 NDKLPLYISNP 721
DB 1187 ntk-----snp 1192
```

RESULT 15

```
AAM60224
ID AAM60224 standard; Protein: 880 AA.
XX
XX AAM60224;
XX
XX 28-SEP-1998 (first entry)
XX
XX Bacillus thuringiensis insecticidal toxin 177C8.
XX
XX Insecticide; pesticide; toxin; delta-endotoxin;
XX biological control; lepidopteran; coleopteran.
XX
XX Bacillus thuringiensis strain PS177C8 (NRRL B-21867).
XX
XX Key Location/Qualifiers
XX FT MISC-difference 253
XX FT MISC-difference 675 /note= "encoded by YRA"
XX FT MISC-difference 846 /note= "encoded by AC"
XX FT MISC-difference 846 /note= "encoded by RAA"
XX
XX W09818932-A2.
```

```
XX 07-MAY-1998.
PD 30-OCT-1997; 97WO-US19804.
XX 30-OCT-1996; 96US-0029848.
XX 30-OCT-1996; 96US-0029848.
XX (MYCO ) MYCOGEN CORP.
XX Dullum CJ, Feltelson JS, Loewer D, Muller-Cohn J;
XX Narva KE, Schmelts JL, Schnepf HE, Schwab G, Stamp L;
XX Stockhoff BA;
XX WPI: 1998-272226/24.
XX N-PSDB: AAV30307.
XX
XX Bacillus thuringiensis isolates - used for producing pesticidal
XX toxins and nucleotide sequences for control of lepidopterans and
XX coleopterans
XX
XX Claim 5; Page 81-84; 139pp; English.
XX
XX This polypeptide comprises a novel soluble toxin of Bacillus
XX thuringiensis (B.t.) strain PS177C8 (NRRL B-21867). The toxin
XX belongs to a novel family of B.t. toxins that have toxicity
XX against non-mammalian pests. Its amino acid sequence was deduced
XX from a novel DNA fragment (see AAV30307) obtained by PCR from
XX cellular genomic DNA of PS177C8. Disclosed and claimed are novel
XX B.t. isolates and toxins (see AAM60218-32) that have activity against
XX lepidopteran and/or coleopteran pests, isolated genes, probes
XX and primers (see AAV30288-321 and AAV99734-87) useful for production
XX of the toxins and for the identification and characterisation of
XX these toxins, and transformed hosts, particularly plant and
XX bacterial hosts. The invention provides 8 entirely new families of
XX toxins from B.t. isolates. The toxins have the additional ability
XX to form pores in cell membranes, and can be used to facilitate
XX entry of a second agent into a target cell.
XX
XX Sequence 880 AA:
XX
XX Query Match 20.8%; Score 781.5; DB 19; Length 880;
XX Best Local Similarity 29.4%; Pred. No. 4,7e-44;
XX Matches 234; Conservative 136; Mismatches 293; Indels 133; Gaps 30;
XX
XX 1 MKRKVLIPLMAISTILVSS--TGNLEVIQAEVK-----OENRLNSESSESSOGLLYG 51
XX 1 mkkk---lasvvtcltlapmfingvnayadskntqstgtknq---qkmdrkgllgy 54
XX
XX 52 YPSDLNFOAPMVTYSTTGDSLIPSELEN--IPSENOYFOSAINSGFTKVKSDPYTRA 109
XX 55 yfkgydf-snlmfaptrdstllydgtanklklkqgqgsrlvlglsksetgdfth 113
XX
XX 110 TSADNHVTMMWDQEVINKASNSNKRIRLEKGRLYOIKIOYOBENPTREKLD-----FKL 163
XX 114 lsedegallieingklisnqkqkqvnhlekqkvlplkleygsd--tkfidskctkelk 171
XX
XX 164 WYTDSONKKEVISDNLOJPELKOKS-----NSRKRKSTAGPTVPDRDND 210
XX 172 fklidqngqgvgqgdellrpefnkkesqeflakpsklnlftqkxmrelded---tdtdgd 228
XX
XX 211 GIPDSLEVEGYTVDVKNKRTPLSPWISNIEKKGLTKYKSSPEKMSSTASDPYSDPEKVTG 270
XX 229 slpdlweengytl-----qnrliavkwddsl-askytklfvsnplshstlvgydpytdyekaar 283
XX
XX 271 RIDKANSPEARHPVAAAYPIVHVDMENILTSKNEDOSTONTSERTISKNTSTSTHTS 330
XX 284 dldlsnaketlnplvaafpsvvnsmekvllspnenls----- 320
XX
XX 331 EVHGNAEVHANTSTSRTHSEVHGNAEVH-----AAVHSHSLAGEERTWAETMG-- 380
XX 321 ---nveshststnwsytnlegasveaglsfgsvnyghsetvage--wgstsgnt 375
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:55:15 ; Search time 132.69 Seconds
(without alignments)
615.118 Million cell updates/sec

Title: US-09-747-521-4_COPY_178_735

Perfect score: 2871
Sequence: 1 DNQLQPELKKQSSNSKKRRS.....LYISNPYKVNYATKENT 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_17:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.yeast:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2755	96.0	764	2	09R0U2
2	2751	95.8	764	2	09KH69
3	2743	95.5	764	2	09F5R7
4	729	25.4	876	2	09KH41
5	728	25.4	876	2	032739
6	718.5	25.0	879	2	006498
7	715.5	24.9	875	2	046221
8	643.5	22.4	721	2	086171
9	211.5	7.4	204	2	09X377
10	181.5	6.3	2178	2	046149
11	166	5.8	604	5	026021
12	164.5	5.7	2529	2	025579
13	158	5.5	4688	2	09P008
14	157	5.5	1308	5	096129
15	155.5	5.4	559	5	09U3Y8
16	150.5	5.2	2269	5	026223
17	150.5	5.2	2399	2	09ZKS9
18	150	5.2	1127	12	09YVT6
19	150	5.2	4919	2	09ZHL0

20	149.5	5.2	2747	5	09BUX9	09bux9 plasmidium
21	149	5.2	1193	2	P71107	P71107 clostridium
22	149	5.2	1193	2	045914	045914 clostridium
23	148.5	5.2	622	2	09U0K4	09u0k4 plasmidium
24	148	5.2	810	2	09CFD3	09cfd3 lactococcus
25	147	5.1	1156	5	077317	077317 plasmidium
26	147	5.1	1939	5	025662	025662 plasmidium
27	146.5	5.1	655	5	026109	026109 plasmidium
28	146.5	5.1	3978	5	097236	097236 plasmidium
29	145	5.1	2647	5	09U4X0	09u4x0 plasmidium
30	144.5	5.0	7107	5	09AF7	09af7 drosophila
31	144	5.0	2402	2	09AER7	09aer7 staphylococ
32	144	5.0	2522	5	077365	077365 plasmidium
33	143.5	5.0	1786	5	09U0P0	09u0p0 plasmidium
34	143	5.0	2340	2	09ZD91	09zd91 rickettsia
35	142.5	5.0	1072	2	09CF64	09cf64 lactococcus
36	142.5	5.0	6658	5	076281	076281 drosophila
37	142	4.9	1595	2	052373	052373 caldicellul
38	141.5	4.9	1893	5	09NKC9	09nkc9 drosophila
39	141.5	4.9	1928	5	09U0H2	09u0h2 plasmidium
40	141.5	4.9	4152	2	09ZHL3	09zhl3 haemophilus
41	141	4.9	712	13	09S8N5	09s8n5 gallus gall
42	141	4.9	713	2	09CM95	09cm95 pasteurilla
43	141	4.9	786	2	09FDM5	09fdm5 streptococc
44	141	4.9	2771	5	026216	026216 plasmidium
45	140.5	4.9	1557	2	09RNI2	09rni2 haemophilus

ALIGNMENTS

RESULT 1
ID 09R0U2 PRELIMINARY; PRT; 764 AA.
AC 09R0U2:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PX01-110 (PROTECTIVE ANTIGEN).
GN PAG.
OS Bacillus anthracis.
OG Plasmid virulence plasmid PX01, and Plasmid PX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=STERNE; PLASMID=VIRULENCE PLASMID PX01;
RX MEDLINE=9945483; PubMed=10515943;
RA Okinaka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hill K.K.,
RA Keim P., Koehler T.M., Lamke G., Kumano S., Mallion J., Manter D.,
RA Martinez Y., Rieke D., Svensson R., Jackson P.J.,
RT "Sequence and organization of PX01, the large Bacillus anthracis
RT plasmid harboring the Anthrax toxin genes.";
RL J. Bacteriol. 181:6509-6515(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=33, 28, AND BA1035; PLASMID=PX01;
RX MEDLINE=99214082; PubMed=10197996;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.,
RT "Genetic diversity in the protective antigen gene of Bacillus
RT anthracis.";
RL J. Bacteriol. 181:2358-2362(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=33, 28, AND BA1035; PLASMID=PX01;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.,
RT Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
RL EMBL; AF065404; AAD32414.1; -
DR EMBL; AF306781; AAG24449.1; -
DR EMBL; AF306778; AAG24446.1; -
DR EMBL; AF306779; AAG24447.1; -
DR EMBL; AF306780; AAG24448.1; -

DR HSP: P13423; IACC.
DR InterPro: IPR003896; Binary_toxB.
DR PRINTS: PRO1391; BINARYTOXINB.
KW Plasmid.
SQ SEQUENCE 764 AA; 85810 MW; 3AE1EBF48FAA03F CRC64;

Query Match 96.0%; Score 2755; DB 2; Length 764;
Best Local Similarity 96.6%; Pred. No. 2, 9e-157;
Matches 539; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

QY 1 DNLQPELQKSSNSRKRSTAGPTVPDRDNDGIPDSLEVEGYTVDAVKNKRTFLSPWIS 60
DB 178 DNLQPELQKSSNSRKRSTAGPTVPDRDNDGIPDSLEVEGYTVDAVKNKRTFLSPWIS 237
QY 61 NIHEKGLTKYKSSPEKMWSTADPYSDPEKVTGRIDKNVSPKARHPLVAAYPIVHDMEN 120
DB 238 NIHEKGLTKYKSSPEKMWSTADPYSDPEKVTGRIDKNVSPKARHPLVAAYPIVHDMEN 297
QY 121 ILSKNEOSTONTDSETRTISKNTSTRTHTSEVHGNAEVANSTSTRTHTSEVHGNAE 180
DB 298 ILSKNEOSTONTDSETRTISKNTSTRTHTSEVHGNAEVANSTSTRTHTSEVHGNAE 357
QY 181 VHAVALDHSLSLAGEPTMAETMGLNADTARLANIRYVNTGAPLYNVLPPTSLVLGKN 240
DB 358 SSTVALDHSLSLAGEPTMAETMGLNADTARLANIRYVNTGAPLYNVLPPTSLVLGKN 417
QY 241 QTLATIKAKENQSLQILAPNNYPSKNLAPIALMAODDFSPPTITNNYNOFLELEKTOL 300
DB 418 QTLATIKAKENQSLQILAPNNYPSKNLAPIALMAODDFSPPTITNNYNOFLELEKTOL 477
QY 301 RLDTQOVYGNATYVNEGRVVDGSGNMSSEVLPQIOETTARILFNGKDLNVERRIAY 360
DB 478 RLDTQOVYGNATYVNEGRVVDGSGNMSSEVLPQIOETTARILFNGKDLNVERRIAY 537
QY 361 NPSDPLETTKPMPTLKEALKIAFGFNEPNGNLOYGCKDTEPDFNDOOTSNIKQOLAE 420
DB 538 NPSDPLETTKPMPTLKEALKIAFGFNEPNGNLOYGCKDTEPDFNDOOTSNIKQOLAE 597
QY 421 LNATNIYTVLDRIKLAKANNILIRDRKFHYDRNNIAVGADESVEKAREVINSSTEGLL 480
DB 598 LNATNIYTVLDRIKLAKANNILIRDRKFHYDRNNIAVGADESVEKAREVINSSTEGLL 657
QY 481 LNIIDDIRKILSGYIEIDTEGLKEVINDRYDMLNISSLRDQKTFIDFKKYNDKPLX 540
DB 658 LNIIDDIRKILSGYIEIDTEGLKEVINDRYDMLNISSLRDQKTFIDFKKYNDKPLX 717
QY 541 ISNPYKVVAVYTKENT 558
DB 718 ISNPYKVVAVYTKENT 735

RESULT 2
Q9KH69 PRELIMINARY: PRT: 764 AA.
AC Q9KH69:
DB 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PROTECTIVE ANTIGEN.
GN PACA OR PAG.
OS Bacillus anthracis.
OG Plasmid pX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V770-NP1-R, ATCC14185; PLASMID=PX01;
RX MEDLINE=20359347; PubMed=10899854;
RA Cohen S., Mendelson I., Alboun Z., Kobiler D., Elhanany E., Bino T.,
RA Leitner M., Inbar I., Rosenberg H., Gozes Y., Barak R., Fisher M.,
RA Kromann C., Velan B., Shafferman A.;

RT "Attenuated nontoxicogenic and nonencapsulated recombinant Bacillus
RT anthracis spore vaccines protect against anthrax."
RT Infect. Immun. 68:4549-4558(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC PLASMID=PX01;
RX MEDLINE=99214082; PubMed=10197996;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RT "Genetic diversity in the protective antigen gene of Bacillus
RT anthracis";
RL J. Bacteriol. 181:2358-2362(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC PLASMID=PX01;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF268967; AAF6457.1; -
DR EMBL: AF306782; AAG2450.1; -
DR InterPro: IPR003896; Binary_toxB.
DR PRINTS: PRO1391; BINARYTOXINB.
KW Plasmid.
SQ SEQUENCE 764 AA; 85838 MW; B0DAFC1DCF83DAF4 CRC64;

Query Match 95.8%; Score 2751; DB 2; Length 764;
Best Local Similarity 96.4%; Pred. No. 5, 1e-157;
Matches 538; Conservative 3; Mismatches 17; Indels 0; Gaps 0;

QY 1 DNLQPELQKSSNSRKRSTAGPTVPDRDNDGIPDSLEVEGYTVDAVKNKRTFLSPWIS 60
DB 178 DNLQPELQKSSNSRKRSTAGPTVPDRDNDGIPDSLEVEGYTVDAVKNKRTFLSPWIS 237
QY 61 NIHEKGLTKYKSSPEKMWSTADPYSDPEKVTGRIDKNVSPKARHPLVAAYPIVHDMEN 120
DB 238 NIHEKGLTKYKSSPEKMWSTADPYSDPEKVTGRIDKNVSPKARHPLVAAYPIVHDMEN 297
QY 121 ILSKNEOSTONTDSETRTISKNTSTRTHTSEVHGNAEVANSTSTRTHTSEVHGNAE 180
DB 298 ILSKNEOSTONTDSETRTISKNTSTRTHTSEVHGNAEVANSTSTRTHTSEVHGNAE 357
QY 181 VHAVALDHSLSLAGEPTMAETMGLNADTARLANIRYVNTGAPLYNVLPPTSLVLGKN 240
DB 358 SSTVALDHSLSLAGEPTMAETMGLNADTARLANIRYVNTGAPLYNVLPPTSLVLGKN 417
QY 241 QTLATIKAKENQSLQILAPNNYPSKNLAPIALMAODDFSPPTITNNYNOFLELEKTOL 300
DB 418 QTLATIKAKENQSLQILAPNNYPSKNLAPIALMAODDFSPPTITNNYNOFLELEKTOL 477
QY 301 RLDTQOVYGNATYVNEGRVVDGSGNMSSEVLPQIOETTARILFNGKDLNVERRIAY 360
DB 478 RLDTQOVYGNATYVNEGRVVDGSGNMSSEVLPQIOETTARILFNGKDLNVERRIAY 537
QY 361 NPSDPLETTKPMPTLKEALKIAFGFNEPNGNLOYGCKDTEPDFNDOOTSNIKQOLAE 420
DB 538 NPSDPLETTKPMPTLKEALKIAFGFNEPNGNLOYGCKDTEPDFNDOOTSNIKQOLAE 597
QY 421 LNATNIYTVLDRIKLAKANNILIRDRKFHYDRNNIAVGADESVEKAREVINSSTEGLL 480
DB 598 LNATNIYTVLDRIKLAKANNILIRDRKFHYDRNNIAVGADESVEKAREVINSSTEGLL 657
QY 481 LNIIDDIRKILSGYIEIDTEGLKEVINDRYDMLNISSLRDQKTFIDFKKYNDKPLX 540
DB 658 LNIIDDIRKILSGYIEIDTEGLKEVINDRYDMLNISSLRDQKTFIDFKKYNDKPLX 717
QY 541 ISNPYKVVAVYTKENT 558
DB 718 ISNPYKVVAVYTKENT 735

RESULT 3
Q9F5R7 PRELIMINARY: PRT: 764 AA.
AC Q9F5R7:
DB 09F5R7:

DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PROTECTIVE ANTIGEN.
OS Bacillus anthracis.
OC Plasmid pX01.
OC Bacteria: Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BAL024;
RA MEDLINE=99214082; PubMed=10197996;
RT Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RT "Genetic diversity in the protective antigen gene of Bacillus
RT anthracis";
RL J. Bacteriol. 181:2358-2362(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BAL024;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF306783; AAG24451.1; -
DR InterPro: IPR003896; Binary toxB.
DR PRINTS: PR01391; BINARYTOXINB.
KW Plasmid.
SQ SEQUENCE 764 AA; 85828 MW; A1845CE1FEDCD93A CRC64;

Query Match 95.5%; Score 2743; DB 2; Length 764;
Best Local Similarity 96.2%; Pred. No. 1.5e-156;
Matches 537; Conservative 3; Mismatches 18; Indels 0; Gaps 0;

OY 1 DNLQPELKOKSSNSRRKSTSGAPTPVDRDNDGIPDSLEVEGYTVDKKRFPLSPWIS 60
DB 178 DNLQPELKOKSSNSRRKSTSGAPTPVDRDNDGIPDSLEVEGYTVDKKRFPLSPWIS 237
OY 61 NIEHKKLTYYKSSPEKWSASDPYSDFEYVGTGIDKNVSEARHPVLAAPYIYHVME 120
DB 238 NIEHKKLTYYKSSPEKWSASDPYSDFEYVGTGIDKNVSEARHPVLAAPYIYHVME 297
OY 121 IILSKNEDOSTONTDESEPTISKNTSTSRTHSEVHGNAEYHANTSTSRTHSEVHGNAE 180
DB 298 IILSKNEDOSTONTDESEPTISKNTSTSRTHSEVHGNAEYHANTSTSRTHSEVHGNAE 357
OY 181 VHAVALDHSLSLAGESTWAEFTMGINTADARLANANIRYVNTGAPYIYNVLPTTSLVLGKN 240
DB 358 SSVYVAIDHSLSLAGESTWAEFTMGINTADARLANANIRYVNTGAPYIYNVLPTTSLVLGKN 417
OY 241 QTLATITAKENQSLQILAPNNYPSKMLAPALNAODFSTPTTMYNOFLELEKTKOL 300
DB 418 QTLATITAKENQSLQILAPNNYPSKMLAPALNAODFSTPTTMYNOFLELEKTKOL 477
OY 301 RLDTQVYGNATYNFENGVRVDTGSMSEVLPOIQTETARIIFNGKDLVVERRTAAV 360
DB 478 RLDTQVYGNATYNFENGVRVDTGSMSEVLPOIQTETARIIFNGKDLVVERRTAAV 537
OY 361 NPSDPLETTKPDMTLKEALKIAGFENPENGNYOGKDIETFDENFDQOSTQNIKNOLAE 420
DB 538 NPSDPLETTKPDMTLKEALKIAGFENPENGNYOGKDIETFDENFDQOSTQNIKNOLAE 597
OY 421 LANTNTITVYDRTIKLANKNMILIRDKRFHYDRNNIAVGADESIVYKHAHREYINSTGGL 480
DB 598 LANTNTITVYDRTIKLANKNMILIRDKRFHYDRNNIAVGADESIVYKHAHREYINSTGGL 657
OY 481 LNTIDKIRKLISGYIVEIEPTGKLEVINDRYDMLNITSSLRQDGTFLDEKKYNDKPL 540
DB 658 LNTIDKIRKLISGYIVEIEPTGKLEVINDRYDMLNITSSLRQDGTFLDEKKYNDKPL 717
OY 541 ISNPNYVNVYAVTKENT 558
DB 718 ISNPNYVNVYAVTKENT 735

RESULT 4
O9KH41 PRELIMINARY; PRT; 876 AA.
AC O9KH41
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CDTB.
OS Clostridium difficile.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCUG 20309;
RA Chang S.Y., Song K.P.;
RT "ADP-ribosylating Binary Toxin Genes of Clostridium difficile strain
RT CCUG 20309";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF271719; AAF81761.1; -
DR InterPro: IPR003896; Binary toxB.
DR PRINTS: PR01391; BINARYTOXINB.
SQ SEQUENCE 876 AA; 98792 MW; 360D62F352E745A5 CRC64;

Query Match 25.4%; Score 729; DB 2; Length 876;
Best Local Similarity 35.3%; Pred. No. 1.1e-35;
Matches 200; Conservative 95; Mismatches 194; Indels 78; Gaps 22;

OY 29 DRDNDGIPDSLEVEGYTVDKKRFPLSPWISNIEHKKLTYYKSSPEKWSASDPYSDF 88
DB 220 DTDNDGIPDSLEVEGYTVDKKRFPLSPWISNIEHKKLTYYKSSPEKWSASDPYSDF 274
OY 89 EKYVGRIDKNVSEARHPVLAAPYIYHVMEIILSKNEDOSTONTSEPTISKNTS 148
DB 275 EKASGSEFDKATKTEARDELVAAPYIVGMEKLLITSTENAST----DQKTVSRATINS 330
OY 149 RTHTSEVHGNAE-----HANTSTSRTHSEVHGNAEVAVALDHSLSLAGESTWAE 202
DB 331 KTEENFTAGVSVNVGYONGFTANVTNNSHITD-----NSTAVQDS--NGE-SWMTGL 379
OY 203 GLNTADTARLANANIRYVNTGAPYIYNVLPTTSLVLGKNQTLATITAKENQSLQILAPNNY 262
DB 380 SINGESAYIANANRYVNTGAPYIYNVLPTTSLVLGKNQTLATITAKENQSLQILAPNNY 438
OY 263 YPSKNLAPALNAODFSTPTTMYNOFLELEKTKOLRLDTQVYGNATYNFENGVRV 322
DB 439 YPKKGLSPALANTMDQSSRLPIPTDQKLDKGQKQIKLETTVOYSGFGTKN-SSQOI- 496
OY 323 VDTGSMSEVLPOIQTETARIIFNGKDLVVERRTAAVNSDPLETTKPDMTLKEALKIA 382
DB 497 VTEGNSMSDYISQIDISASITLITDEN-ESYERRVTAKNLQDDPDCKT-PELTIGEAIEKA 554
OY 383 GFENPEPKNLOYOQKDIETFDENFDQOSTQNIKNOLAEIANTNTITVYDRTIKLANKN 440
DB 555 FGATKKDGLVFNFDIPIDESCVELIFDNTANKIKDSKTLSDKKIYNY--KLERGMN 610
OY 441 ILIRDKRPFHY--DRNNIANGADESVYKHAHREYINSTGGL-----LNLIDDIR 488
DB 611 ILIRDKRPFHY--DRNNIANGADESVYKHAHREYINSTGGL-----LNLIDDIR 668
OY 489 KILSGY-----IVEI-----EDTEGL-----KEVINDRYDMLNIS-SLRQD 523
DB 669 YVFGYSGKDPFLTSIIVKIKAKKEKTDYLVPEQGYTKFVSFEPTTEKSSNIEFTLIGS 728
OY 524 GKTFIDRKKYND--KLPLYISNPNYKV 548
DB 729 GTTYLDNLSTIELNSTPEILDEPEVKI 755

RESULT	5			
032739				
ID	032739	PRELIMINARY:	PRT:	876 AA.
AC	032739:			
DT	01-JAN-1998	(TREMBLrel. 05, Created)		
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)		
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)		
DE	ADP-RIBOSYLTRANSFERASE.			
CN	CDTB.			
OS	Clostridium difficile.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_Taxid=1496;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDI96;			
RX	MEDLINE=97230316; PubMed=9119480;			
RA	Perelle S., Gilbert M., Bourlioux P., Corthier G., Popoff M.R.;			
RT	"Production of a complete binary toxin (actin-specific ADP-			
RL	ribosyltransferase) by Clostridium difficile CD196.";			
RL	Infect. Immun. 65:1402-1407(1997).			
DR	EMBL: L76081; AAB67305.1; -.			
DR	HSSP: P13423; 1ACC.			
DR	InterPro: IPR003896; Binary_toxB.			
DR	PRINTS: PRO1391; BINARTOXINB.			
KW	TRANSFERASE			
SQ	SEQUENCE 876 AA: 98796 MW: 2506E2D45CE2B3B CRC64:			

Query Match	25.4%;	Score 728;	DB 2;	Length 876;
Best Local Similarity	35.3%;	Pred. NO. 1.2e-35;		
Matches 200; Conservative	95;	Mismatches 194;	Indels 78;	Gaps 22;

QY	29	DRDNGIDPSLSEVEZYADVCKRRFELSPWISINIEHKCLTYKSSPEKMSVSPADISYDF	88
Dd	220	DTDNDNIDDSYRNKGT-----KDLIAKMWESFPE-QCYKRYVSYNLESNAGDPYDY	274
QY	89	EKYVTRIDKNVSPEARHPLVAAYPIVHVDMENILSKNEDOSTONTDEPRTISKNTSTS	148
Dd	275	EKASSFPKAITETARDPLVAAYPIVGVOMEKLITSTNEHAST---DQKTVSRATNTS	330
QY	149	RTHSEVGNALVE-----HANTSTRKTHSEVHGNAEYHAVAIDHSLSLAGEFTVAETM	202
Dd	331	KTESNTAGVSVMVGONGEFTAVNTVNTSHHTD-----NSTAVODS---NGE-SWNTGL	379
QY	203	GLNTADTARLANINRYVNTGTAPITNVLTSLVGLKNOTLATIKAKENOQIILANPNY	262
Dd	380	SINKESAYIANNAVRYNTGTAPMKVPTPTMLVL-DGDTLESTIKAOEQIGNNLSPGOT	438
QY	263	YPSKNLAPIALNAODFSSPTPTMNYNOFLELEKTRKOLRLDPOVYGNATATYFENGVRV	322
Dd	439	YPKKLSPLALNTMTOFSSRLIPINDQKLKADAGQIKLETTQVSGNNGTKN-SSGOL	496
QY	323	VDTGNSMSEVLPOIOETARITIFNGKDLNVRRRIAANVSPPLETTKTKDMFLKEALKA	382
Dd	497	VTEGMSWSDYISQDISAISILIDPEN-ESYRRRYTAKMLQDPEKT-BELTIGEAIEKA	554
QY	383	FGFNPNGNLQYOGKDITE--FDPNPDQOSTONIKNOALENATNATYVLDKIKKANKN	440
Dd	555	FGATKKDGLIYNNDPIDESCYELLFPDONTAKIKDSLTKLTSDDKTIYV---KLERGMN	610
QY	441	ILIRDKRHHY---DRNNIAVGADESVMKAREVINSSTEGL-----LLNIDKDIR	488
Dd	611	ILIKRPTVFTNDDOYNNP--STWSNVNNTTNDGIGSANKLNGETKIKIIPSELKPYKR	668
QY	489	KILSGY-----IYEI---EDTEGL-----KEYINDRYMLNIS-SLRD	523
Dd	669	YVFSSYKSDPLTNSIIVYIKAKERKTDVLVPEOGYTKRSEYFETTEKSDSNIEITLLGS	728
QY	524	GKTFIDFKKYN-D-KLPLYISNPNKV	548
Dd	729	GTTYIDNLSITELNSTPELDEPEVKI	755

RESULT	6			
006498				
ID	006498	PRELIMINARY:	PRT:	879 AA.
AC	006498:			
DT	01-JUL-1997 (TREMBLrel. 04, Created)			
BT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	SB COMPONENT.			
GN	SBS.			
OS	Clostridium spiroforme.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes.			
OX	NCBI_TaxID:29348;			
BN	[1]			
RP	SEQUENCE FROM N.A.			
EC	STRAIN-CS246;			
EA	Gilbert M., Perelle S., Daube G., Popoff M.R.;			
EL	Syst. Appl. Microbiol. 20:337-347(1997).			
DR	EMBL; X97969; CAA66612.1; -.			
DR	HSSP; P13423; IAC.			
DR	InterPro; IPR003896; Binary_toxB.			
DR	PRINTS; PRO1391; BINARYTOXINB.			
SO	SEQUENCE 879 AA; 98738 MW; 40685ACBBE05BA01 CRC64;			

Query Match	25.0%;	Score 718.5;	DB 2;	Length 879;
Best Local Similarity	29.5%;	Pred. No. 4.6e-35;		
Matches 209;	Conservative 115;	Mismatches 187;	Indels 197;	Gaps 27;

0y	7	ELKSSSRKKRSTASAPYDPBNDGIPDSELEGEYDVYKKNRTFLSPISINHEKK	66
0b	203	DUKLSRSAR--LASGWDDEDLDIDNDNDIPDYENKNGTIT---KDSIAVKN-EDSFAQ	255
0y	67	GLTRYKSSPEKMWSTADSEYDFEKTVTGRIDKNVSPEARHPLVAAPYIVHDEMILSKN	126
0b	256	GYYKLLSYLESNTAGDPDYQKASGSPDKAIKAEARDPVAAAPYVGVGHEKILISTN	315
0y	127	EDQSTQNDSEKRTISKNTSRT--HISEVHGNAEVH---ANSTSTKTHSEVHGNAE	180
0b	316	EHASZ---DQKIVSRNTTNSKTSANAGVAINIAYONGFTGSLTTNYSHTTE-----	365
0y	181	VHAVYDHSLSLAGERTAETGMLTADTAPRANANIRYVNTGTAPILYVLPPTSVLGKN	240
0b	366	NSYAVQNS---NGE-SMNTSLSIKKGESAYINNAVRYRNTGTAMRYKTPPTNIVL-DG	419
0y	241	QTLATIRAKENOLSIAPNNYPSKNIAPIALNADODESSFTPTMNTYQPLELEKTOL	300
0b	420	DTLLTIIKADNOIGNNLSPENETYPKKGSLPLIINTMDPFSFLPIINLYDQLEKLGAQOI	479
0y	301	RLDPDQVYGNLATYFENGVRVDP--GSWMSVLVLPJOETTRRIIFN-GKDLNVERRA	358
0b	480	KLEITVOVGN--YGIKNSOGOLITEGNSWSYISOLDSLSIILDGSD--VEERKAT	534
0y	359	AVNPSDPLETTRKPDMTLIEALKIAEGFNEPNCNLOYOGKDIYE--FDNFDOOTSINIK	416
0b	535	AKDSSNPEDKT-PVLTIGEAIEKARGATKNGEILYFNGMPIDESCVELIFDGNFANLIKE	593
0y	417	QLAELNATINITYVDKIKLNAKMLIIRD-----	445
0b	594	RLNALNDKKIYVNV---OLERGMKILITITSTYFNNFPGYNNPDSWSMVSNDNODLONA	649
0y	446	-----KR	447
0b	650	ANKLSGEFKIYIPMKLNPYKRYVSGYLKNSSTSNPTITVNIKAKEQKTYNLVSENDYK	709
0y	448	FHY-----DRNN-----IAYGADESVYKEAH	468
0b	710	FSYEPETIGRASNIEITLTSQTFILDNLSITELNSTPEILKEPDIKVPDOEII-DAH	768
0y	469	REV-----INSTEBLLN-----IDKDIRKILISGIVEIE--DTEBLEKVINDRYDM	514
0b	769	KYYVADISFNOSTANVYLDGLFEPTQTNKKELDVYQKRYKATLEYSGFKQIGTRKDEL	828

```

QY      515 LNISSLRODGT-FIDFKKYNDKLPYISNPY----KVNVAATKEN 557
      | : ||:::| : | : :|||
Db      829 RNYTGDSNQPKTNYVNFRRSY-----FTSGENVMPYKKLRITAITPEN 870

```

RESULT	7		
046221			
ID	046221	PRELIMINARY:	PRT: 875 AA.
AC	046221;		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	IOTA TOXIN COMPONENT IB PRECURSOR.		
OS	Clostridium perfringens		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;		
OC	Clostridium.		
OX	NCBI_TaxID=1502;		
PN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-NCIB 10748;		
EX	MEDLINE-94041637; PubMed-8225592;		
SA	Pelletle S., Gilbert M., Boquet P., Popoff M.R.;		
RT	"Characterization of Clostridium perfringens Iota-toxin genes and		
RL	expression in Escherichia coli.";		
RL	Infect. Immun. 61:5147-5156(1993).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-NCIB 10748;		
SA	Popoff M.R.;		
RL	Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-NCIB 10748;		
SA	Popoff M.R.;		
RL	Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.		
RL	Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.		
SA	Popoff M.R.;		
DR	EMBL: X73562; CAAS1960.1; -.		
DR	HSSP: P13423; IACC.		
DR	InterPro: IPR003896; Binary_toxb.		
DR	PRINTS: PR01391; BINARYTOXINB.		
KW	Signal.		
FT	SIGNAL		
ET	CHAIN	34	38
ET	CHAIN	212	875
EQ	SEQUENCE 875 AA; 98468 MW; C9A9E092CD3818921 CRC64;		
			POTENTIAL.
			IOTA TOXIN COMPONENT IB.

Query Match	24.9%	Score 715.5	DB 2	Length 875
Best Local Similarly	32.6%	Pred. No. 6.9e35		
Matches 205	Conservative 102	Mismatches 207	Indels 115	Gaps 26

[illegible][illegible]

```

RESULT      8
086171
ID      086171      PRELIMINARY;      PRN;      721 AA.
AC      086171;
DT      01-NOV-1998 (TREMBLrel. 08, Created)
DT      01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT      01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE      C2 TOXIN (COMPONENT-II).
OS      Clostridium botulinum.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC      Clostridium.
OX      NCBI_TaxID=1491;
RN      [1].
RN      SEQUENCE FROM N.A.
RC      STRAIN=TYPE C (C) -203U28;
RX      MEDLINE=98323874; PubMed=9659689;
RA      Kimura K., Kubota T., Ohishi I., Isogai H., Isogai E., Fujii N.;
RT      "The gene for component-II of botulinum C2 toxin.";
RL      Vet. Microbiol. 62:27-34(1998).
RN      [2].
RN      SEQUENCE FROM N.A.
RP      STRAIN=TYPE C (C) -203U28;
RC      STRAIN=TYPE C (C) -203U28;
RX      MEDLINE=96184657; PubMed=8645309;
RA      Fujii N., Kubota T., Shirakawa S., Kimura K., Ohishi I., Morishita K.,
RT      Isogai E., Isogai H.;
RT      "Characterization of component-I gene of botulinum C2 toxin and PCR
RT      detection of its gene in clostridial species.";
RL      Biochem. Biophys. Res. Commun. 220:353-359(1996).
DR      HSPD: D84982; BAA32537.1; -.
DR      HSSP: P13423; IACC
DR      InterPro: IPR003896; Binary_toxB.
DR      PRINTS: PR01591; BINARYTOXINB.
SO      SEQUENCE 721 AA; 80515 MW; 44C8153AC749D5F2 CRC64;

```

Query Match	22.4%	Score 643.5	DB 2	Length 721
Best Local Similarity	35.5%	Pred. No. 1.1e-30		
Matches 1/1	Conservative 83	Mismatches 161	Indels 67	Gaps 16

```

0Y      8 LKONSSNRKKRSTASGPTV-----PBDNDNGIPSLVEGYTDVKKRRELS 56
Db      155 LKPNYSNTNEKSKFIPNPTLSSNAKLLKANANRBDROGIPREWLINGTYNNOKAVAMD 214
0Y      57 PWISNIEHKCLITKYKSSPERKMTASDPYDFEKYGTGRIDKNSPEARHPLVAAVPIVH 116
Db      215 KFAAN-----GYKTVSNPFKPCZANDPPTDFEKYSGQIDPSSMAADPMSATPIVGV 263
0Y      117 DMENIIISKNDQSTONTDSETRTISKNTSTSRTH-----SEVGNANE----- 161
Db      270 QMERIVYVSKSE---TIIGDSTKMSKSTSHSSNINMYGAEVSGSLQDLNGGIPFPVSM 325
0Y      162 -HANTSRTHTSEVHGNAEYHAVAIDHSLSLAGERTWELMTGINTADTARLANIRYN 220

```

```
Db 336 ASANSHWOMSTYVD-----DTTGE-SFSQGSINTGESAYINPIRYN 370
Oy 221 TGTABIYVNLPTTSVLGKNOTLATIKAKENQISOILAPNNYPSKNLAPINQODDS 280
Db 371 TGTAVYVNTPTTTIVIDK-QSVAITIKQESLISGYLNGCGYPIIGEPMALNTMDOFS 429
Oy 201 SPTPTMNTNOFLEKTKQLRLDQOYGNATYVNFENGVRVDTGSMSEVLPQIOETT 340
Db 430 SRLIPINYNQLKSINGCGYVLMSTQSGFNFAKYN-SNGNLVTD-GNMNGPYLGTIKSTT 487
Oy 341 ARII--FNGKDLNVERIAVNPSPDPLETTKPDMTLKEALIAFGFNPNGNLQYOGKD 398
Db 488 ASLTSESGQTTQVA---VVAPESDPEDKT-PKLTLQALVKAFALEKKNGKEYFHGLE 543
Oy 399 IT---EPDFNPDQOTSQNIKNQALAEIENATYVLDKIKLNKNNILIRDKRFHYDRNNI 455
Db 544 ISKNKIKQVFLDSNTNPNFENQLNKTADKIMHCT--IKRN--NMILVAVITFKENISSI 599
Oy 456 AV 457
Db 600 NI 601

RESULT 9
Oy 09X377 PRELIMINARY: PRF: 204 AA.
Oy 09X377
AC 09X377:
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE PX01-111.
OS Bacillus anthracis.
OC Plasmid virulence plasmid PX01.
OC Bacillus/Firmicutes: Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=STERNE;
RA Okinaka R.T., Cloud K., Hamton O., Hoffmaster A., Hill K.K., Keim P.,
RA Koshler T., Lamke G., Kumano S., Mahillon J., Manter D., Martinez Y.,
RA Rieke D.O., Svensson R., Jackson P.J.;
RT "The sequence and organization of px01, the large Bacillus anthracis
RT plasmid harboring the Anthrax toxin genes.";
RL J. Bacteriol. 0:0-0(1999).
DR EMBL; AF065404; AAD32415.1; -.
DR HSSP; P13423; IACC.
KW Plasmid.
SQ SEQUENCE 204 AA; 23029 MW; E1657B23AE4273FD CRC64;

Query Match 7.4%; Score 211.5; DB 2; Length 204;
Best Local Similarity 34.8%; Pred. No. 1.5e-05;
Matches 46; Conservative 31; Mismatches 34; Indels 21; Gaps 4;
Oy 439 ENILRDKRFHYDRNNIAGADESVYKAEHREYINSSTEGLLNIDKDIRKLSGYIEI 498
Db 1 MNILVPRDE-YHYDNGNINIVGDVSYLKNAAYKQILNMSSDGVSLNDEVDNALSGYMLQI 59
Oy 499 EDTE-----GLKEVINDRYDMLNLSLRQDGKTFIDPKKYNIDKPLIYSNPN 545
Db 60 KKPNSHLTNSPVTTTLAKRDSGVGELYRVL5-----DQTEGLNKEKEDENMRSLV-DPG 112
Oy 546 YKVNYYAVTKEN 557
Db 113 DDVYVYAVTKED 124

RESULT 10
Oy 046149 PRELIMINARY: PRF: 2178 AA.
Ac 046149; 046147; 046148;
```

```
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ALPHA-TOXIN.
OS Clostridium novyi.
OC Bacteria; Firmicutes: Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC19402;
RX MEDLINE=95342160; PubMed=7616958;
RA Hofmann F., Herrmann A., Habermann E., von Eichel-Streiber C.;
RT "Sequencing and analysis of the gene encoding the alpha-toxin of
RT Clostridium novyi proves its homology to toxins A and B of Clostridium
RT difficile.";
RL Mol. Gen. Genet. 247:670-679(1995).
RN [2]
RP SEQUENCE OF 1204-2178 FROM N.A.
RC STRAIN=ATCC19402;
RA Hofmann F., Habermann E., von Eichel-Streiber C.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z48636; CA88565.1; -.
DR EMBL; Z23280; CA80818.1; -.
DR EMBL; Z23281; CA80819.1; -.
DR InterPro: IPR002479; CW_binding.
DR Pfam: PF01473; CW_binding_1; 10.
SQ SEQUENCE 2178 AA; 250134 MW; 9B0ADCE031CA4A75A CRC64;

Query Match 6.3%; Score 181.5; DB 2; Length 2178;
Best Local Similarity 20.7%; Pred. No. 0.026;
Matches 122; Conservative 95; Mismatches 196; Indels 177; Gaps 27;
Oy 63 HKKKGITRYKSSPEKMWASDPYDFEKYTGRIKNSPEAR----HPLVAYPIVHDM 118
Db 320 NEKKLMNNY-----PYKMEQVPSIEKERIISFVNHHIINDILPLGDIKISQ 367
Oy 119 ENILSKNEDOSTONTSETRTISKNTS-TSRTHTSEVHG----NAEVHANTSTSRHT 172
Db 368 LEILLSRLKAATGKKTFSNAFIISNNDSLTLNLLISQLENREYELINSIOEKFKICEYD 427
Oy 173 SEVHGNAEVHAVAIDHSLSLAGEPTAETMG-LNTADARLANANRY-----VNGTAPI 226
Db 428 SYINSVSELVLETTPTKNNLSMDSSSEFYQOIGYLSGFPREVSVTFEGSPNITYSATCDT 487
Oy 227 VVNLPTT-SLVGKNOTLATIKAKENQISOILAPNNYPSKNLADIALNAODPFSTPI- 284
Db 488 YHFKNTPTDMLSSQDET-----FEASNNLYFSK-----THDEFKSSMLL 527
Oy 285 --TMNYNOFLEKTKQLRLDQOYGNATYVNFENG-----VVDGSGSMSEVLPQIO 337
Db 528 RSNIAEKERQKLIKTYIGR-----TLNYEDGLNFMNKKRYTT-----SELKKYIE 572
Oy 338 ETTAATITNGKDLNVERIAAVNPSPDPLETTKPDMTLKEALKIFGENEPGNLQYOGK 397
Db 573 EYNSKTIYENDLNNI-----LQIOGDIDSYESAANNV-FGRN-PKKSILIOGV 618
Oy 398 DITEPDFNPDQ--QTSQNIKNQALAEI-----ANINITY 429
Db 619 DDFAVNFYFENGCIQVSDNINNLISRFNDIKIKTLFLIGHGENVFPKLEGGKTVVDLTYN 678
Oy 430 LDKIKLN-----AKNNIL-----TRDKRFH-----YD 451
Db 679 IKPKLOHLEREGVILKNNKYILNLCGYMPTKPVNDINSTFVGLFNKISRDLPKGS 738
Oy 452 RNNIAGADESVY---KEAHREVINSSTEGLLNIDKDIRKLSGYIV---EIDTEGK 505
Db 739 KNQLEISANKYAIRINREKREVLDYFGK-WVSNDLIAEQISNKYVYVYWNVEVNT---- 793
Oy 506 EYINDRYDMLNLSLRQDGKTFIDPKKYNIDKPLIYSNPNYK---VNYA 552
Db 794 --LSARVEQLN-----KVAEFAKIDINSIIQTNNQELKOSLVNYIA 832
```


RESULT 11
ID 026021 PRELIMINARY: PRT: 604 AA.
AC 026021:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE STAP-ANTIGEN.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=79/96;
RX MEDLINE=95021499; PubMed=7935600;
RA Fildock D.A., Bottius E., Braham K., Moelans I.T.M.D., Alkawa M.,
RA Konings R.N.H., Cerna U., Olafsson P., Kaldoh T., Asavanich A.,
RA Guerlin-Marchand C., Druilhe P.;
RT Cloning and characterization of a novel Plasmodium falciparum
RT sporozoite surface antigen, STAP-1;
RL Mol. Biochem. Parasitol. 64:219-232(1994).
DR EMBL: Z26314; CAA81224.1; -
SQ SEQUENCE 604 AA; 67051 MW; DE23D9442A509667 CRC64;

Query Match 5.8%; Score 166; DB 5; Length 604;
Best Local Similarity 21.3%; Pred. No. 0.037;
Matches 125; Conservative 83; Mismatches 250; Indels 128; Gaps 28;

QY 45 TVYVKKRFFLPWISN-----IHEKKG---LTKKSSSEKSTASD-PYSD 87
DB 39 TVYIKRNR-FLSEYQSNFLGGYSAALKLVNSKSGTNTVTKYNS--ENTNTNNIPSS 95
QY 88 FEKVTIRIDKNVPEARHPVLAAYPIYVDMENILSKNED-----OSTONTOSETRTIS 142
DB 96 SYTYNTRLAANS-----TTSTTKYTDNKKTKIKLGNSTIINTSTENT-SATKKVY 149
QY 143 K-----NTSRTHTSEVHGNAEYHANTSTSRHTSEVHGNAEYHAVAIDHSL 191
DB 150 ENVTNQILTGNNNTTNTSTTEHNNINNTNTSTENTSATKKTEVNTNQILTGNNNT 209
QY 192 LAGEPRWAEHTMGLT---ADTARLANINRYVMGTAP--IYVNLPTSLVLGKQTLAT 245
DB 210 TTNFTSTTEHNNINNTNTNSTDNSNTJNTLT-DNTSTTKKLTLDNINTQNLTTSTNTTYS 268
QY 246 -----IKAKENQLSQILAPNNY-----YPSKNLAPIALNADDPSPITMYNY-QF 291
DB 269 TDNNNTINTKPIDNNNTDIKFTDYNNTGKTFDKKN---TDIKATDNNNTITTTDNTNTNY 325
QY 292 LLEKTKQLRLDTDOYVGN-IATYFNENGRVYDGTSGNMSEVLPOIETTARIIFNKDL 350
DB 326 ISTDNSKTNVISTDNSKTNISTDNDNADTILTDNDNNTDIIILTD-----NNNTDT 376
QY 351 NLVERILAAVNPSPDELTTPDMTKLEALKIAGFNPENGLQYOGKDITEFDNFOQT 410
DB 377 ISTDNDNADTKATDNNNTNTKATDNNNTKI--SPDNN---NRTKSTDDNNNTNT 428
QY 411 SQRIKQALQALNTATYTVYDJKILNAKMIILRDKRFPHDRNNI--AVGADSESVKFAHR 469
DB 429 KATDNNNTKTIISTDNNNT-----KTIISDNNNTKTIISTDNNNTKTIISN 471
QY 470 EVINSSST-----EGLLLN-----IDKDIRKILSGYIEIDTEGLKEVIN--- 509
DB 472 DNNNTMTISTDNNNNMTNQVFPANNYNETSDDDELNKCODY---SEKEIKKSMINAVL 528
QY 510 DRKDMNLISLRDQKTFIDFKKYNDKPLXISPNKYKVAVYATK 555
DB 529 DKLDLETVRKIHSDISTGLE-KKNRP--NOIITHLNLNKKMYNLIK 571

RESULT 12

025579
ID 025579 PRELIMINARY: PRT: 2529 AA.
AC 025579:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE TOXIN-LIKE OUTER MEMBRANE PROTEIN.
GN HP0922.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group.
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.C.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-B., Gill S., Dougherty B.A.,
RA Nelson K., Richardson J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
DR EMBL: AE000602; AAD07969.1; -
DR TIGR: HP0922; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2529 AA; 274561 MW; 440882E8644472EC CRC64;

Query Match 5.7%; Score 164.5; DB 2; Length 2529;
Best Local Similarity 19.4%; Pred. No. 0.033;
Matches 113; Conservative 83; Mismatches 206; Indels 181; Gaps 26;

QY 120 NIIISKN--EDOSTQNTDSETRTSKNTSTSRHTSEVHGNAEYHANTSTSRHTSEVHG 177
DB 825 NIYLTNNFKTGEVSNBDDGGANITPKASDNIITMDGLNYNDAE-----TVTKMIQT 875
QY 178 NAEVHAVAIDHSL-----SLAGERTW-----AETMGLTFAD-----TAR 211
DB 876 GASQSHATPDALNNISYTNSSFSQMTWGRKFSKAKNISFASNAFSGFTNPGSGSVISAN 935
QY 212 LMANIRYVNT--GTAPIYVNLPTSLVLGKQTLATI-----KAKENQLSQILAPNNYYP 264
DB 936 ATNLSFINSRLNGCAYVN-LQANSLIFNNQAVFNVLVSRTGSINPATQTLGNTNFTL 994
QY 265 S-----KNLAPIAL--NAODDSSPTITANYNOFLEKTKQLRLDTDOYVG 309
DB 995 SSQSLNFGNDPTTLQNNANNTILGKNSQAAPFKN--LTLDDNNSLSLDQSVLANNTSAFN 1053
QY 310 NIATYFNENGRVYDGTSGNMSEVLPOIETTARIIFNKDLN-VERRIAAVNPSPDELT 368
DB 1054 NQASLNINYS-----QATFNSLFFNGTSLNASSKLNASNASFSNNT 1097
QY 369 TKPDMTKLEALKIAGFNEPNGLQYOGKDITEF-----DFNPDQOSTONIKQOLA 419
DB 1098 T---INLDDSVLSASNTSLNANINFGASQADGCGNTIITDASPNFDSASLNFNNLTA 1154
QY 420 -----ELANTYTVYDJKILNAKMIILRDKRF- 448
DB 1155 NGALNFGYPSLTALKALMSVSGQFVLGNNGDINLSDI-NIPDNTTKSVTYNILNAOKIT 1213
QY 449 -----HYDRNNIAYGA-----DESVKFAHR-----EVIN- 473
DB 1214 GISGANGYEKLFYGMKIQNATYSDDNNNTQWSTINPLNSQIIQIESIKNDLITVELNN 1273
QY 474 -SSTEGLLNIDKDI-----RKILSGYIEIDTEGLKEVINDRYDMLNITSL- 520
DB 1274 PNSASNTIFNPAPLAYNQASKQNPDTGYSYSDNQA-----GTYVLTLSNKGFLFTPKGS 1328

QY 521 --RODGTFIDFKKYNKLPY---ISNPNTKVVNYAVTKENT 558
DB 1329 QTPQAPGTSPFNQPLSSILNTYKNGFSSSENLK-TLLGILSONS 1370

RESULT 13

09P008 PRELIMINARY; PRT; 4688 AA.
ID 09P008
AC 09P008;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE HYPOHETICAL PROTEIN U0482.
GN U0482.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmatellaceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROVAR 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
urealyticum.";
RL Nature 407:757-762(2000).
DR EMBL: AE002145; AAF30894.1;
DR InterPro: IPR001152; Thymosin_b4.
DR InterPro: IPR000626; Ubiquitin.
DR Pfam: PF00240; ubiquitin; 1.
DR SMART: SM00152; THY; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 4688 AA; 534880 MW; B53ABFAFFEE1997E CRC64;

Query Match 5.5%; Score 158; DB 2; Length 4688;
Best Local Similarity 21.2%; Pred. No. 1.9;
Matches 138; Conservative 107; Mismatches 270; Indels 136; Gaps 32;

QY 1 DNLOPELKOKSSN-----SRKRSFSAGPIVPPDR-----NDGIPD 37
DB 3803 DNLN-PEKRYKLENTLSKPLKTHNLSVINDKENISLITETGPNVLKIQTONDITND 3861
QY 38 SLEVEGYVDVYKNTFLSPWISNIHEKKGILTKYKSSPEKMTASDPYSEKVGRIK 97
DB 3862 TQGTINVLGNSK-YNGRQIKVYKDNNNVITYESS---LITLQKNDYOLLNLNS 3917
QY 98 NVSPARHPVLAAYPIVIV-----DME-----NIIISKNEOSTONTDSE-----TR 139
DB 3918 N-----REYREKEIENIHISNTNNEFELEKNGVSNFTQTKNTYVOMNDSATIVGTR 3972
QY 140 TISKNTSRTHTSEVHGNAEY---HANTSTRTHTSEVHGNAEYHAAVADHLSLAGER 196
DB 3973 GVNREFKI-KSEDKLENNQOVANFAKETIKROTNTMLQITRPLKDYTSDEK-----EG 4026
QY 197 TWAETMGLNT---ADTARLANIRVYNTGTAPIYVNLPTSLVLGKNOTLA-----TIK 247
DB 4027 TWAHLSNVNPKRETTYKLVKIOGVNPKAKKNINNSENNVLIDNINSINSNYEFTK 4086
QY 248 AKENOLSLAPNNYPSKKNLAPALN-AODEFSPTITMAY--NQFLELEKTKQLRDT 304
DB 4087 VGDHLNITSSNNVNTSQTINFTLSCGKKSWMGKIKLISKYSDTSESITHTNEVLIES 4146
QY 305 DOYVGNITYPENGR-----VRVDTSNNSEVLPOIETARTILFNGKDLNVERKIA 358
DB 4147 NKTQYINILNKKRNRYTLIDVKLIDNNNSVDFPEKGNLNSFTTETSAINVIETI 4206
QY 359 AVNPSDPLETT-----KPDWTLKEA-----LKIAFGNPNNGNLQYOGDITEF 402
DB 4207 SNRSTNLSKSTLIKINLNDPDLNLRDKDOATIVYGNKKQAMGFIVSGNIKILTLTAVDL 4266

QY 403 DNFNDOOTSNIK-NQALNATNTIYVLDKIKLAKANNILIROKREHYD-----RNN 454
DB 4267 NFN-DKVINIVISFNKKSISAEN-----IGIDSKNII-----YNNDSIPKLEIND 4313
QY 455 IAYGA---DESVYKKAHEVINSSEGLLINDKIRILSGYIYEIDTGKLEKIND- 510
DB 4314 IIVNGPINKETVYKNAK--NNIDVDGLQINPKIAHNL-R-FAKFKSTN-----NDI 4364
QY 511-RYDMLNITSL-RODGTFIDFKKYNKLPYISNPNTK-VVNYAVTKENT 558
DB 4365 IETVINGSSLVNNDGKTSIRFTLNK-----ANKXLSVDVLYLVNNS 4410

RESULT 14

096129 PRELIMINARY; PRT; 1308 AA.
ID 096129
AC 096129;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE PREDICTED MEMBRANE ASSOCIATED PROTEIN.
GN PFB0125C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettein H., Carnuci D.J., Cummings L.M., Aravind L.,
Koonin E.V., Shalton S., Mason T., Yu K., Fujii C., Pederson J.,
Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RA "Chromosome 2 sequence of the human malaria parasite Plasmodium
falciparum.";
RT Science 282:1126-1132(1998).
RL EMBL: AE001374; AAC1815.1;
DR InterPro: IPR001313; PDM.
SQ SEQUENCE 1308 AA; 155585 MW; 9722F0336606C366 CRC64;

Query Match 5.5%; Score 157; DB 5; Length 1308;
Best Local Similarity 19.9%; Pred. No. 0.37;
Matches 130; Conservative 95; Mismatches 219; Indels 210; Gaps 31;

QY 47 DVKNKRTFLSPWISNIHEKKGILTKYKSSPEKMTASDPY----- 85
DB 42 DVEKKDVTL-----NLDEKKNVEEYKKNKDVFKNEDEFFVEFDKEINKLKIKEGCN 96
QY 86---SDFEYVIGRI--DKNVSPARHPVLAAYPIVHVDMENIILSKNEOSTONTDSETRT 140
DB 97 MKEINFIEKGYILNDEVNS--TINNITSLNNDILHSSDKNCTSYNIPSGNNNNN-- 153
QY 141 ISKNSTSTRTHTSEVHGNAEYHANTSTRTHTSEVHGNAEYHAAVADHLSLAGER 195
DB 154 -----NNNNNVISNNS-----NIFVDSHMEHFDDITDTEFEKID 187
QY 196 TWAETMGLNTADTARLANIRVYNTGTAPIYVNLPTSLVLGKNOTLATIKAKENOLSQ 255
DB 188 QTNFSEFPQNTSFEKKNVNEELMKHTDNT-NI---CDKIIDK-----KKKNNTLS 236
QY 256 ILAPNNYPSKKNLAPALN-AODEFSPTITM--NYNOFLELEKTK-QLRDTDOVYGN 311
DB 237 LIHDNLNFDNLNITYEDNNNKDVSITDLMKKNYK--NFEKNEIDVYVDTSTTFENI 294
QY 312 ATYENENRVRVDGSSNMSEVLPOIET-----ARIIFNGDLNVERKIA- 359
DB 295 NNDNNEKNLMLNMQSDKEILNKKNDTFYINNKFLISENNILLEDKDISFIDKIESN 354
QY 360 -----VNSDPLETT-----PDMWTLKE-----AL 379
DB 355 KCEDYCVANNNNNNNNRNLSDILENAYSKDCESRTINEDRIYNFNEEDKISHDAFDIT 414

```

Oy 380 KIAFENPNPNDIYOGADITFEPFNPDQOSTNIK-----NOLAELATNIYVL-D--K 432
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 415 PSSFKKEEDKNCNEKYQ-----NVEFSKNDLENINIVEDPPSSNSE-EKONFQONCDSN 469
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 433 IKLNKKM--NILIRDKRFH---YDRNNIYAGDAESVAKAEHR-----469
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 470 IWLKKKFEHNHVFERNIEYEPKNYENEND---DQKDVDSSKFEFNNEFMDKKKNY 525
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 470 -EYINSNTEG---LLNIDKDIRKILSGIYEI-----EDTEGLKYEYINDR 511
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 536 DEIIDSQVEGNCADVEEIKDKKEEKT---YFYEYGINYGDELPRRNFEEIDENKREYEEK 582
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 512 YDMLNISLRODDKTFID--FKKYNDLPLTYI-----SNPNKYVAVATKE 556
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 563 FD-----EKMGENFEEVEEKEIKDEYVGNIFEVEEKEKDKRENIIFEIE 629
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

[illegible]

Search completed: December 2, 2001, 13:55:20
Job time: 555 sec

	RESULT	15
09U3Y8		
ID	09U3Y8	PRELIMINARY; PRT; 559 AA.
AC	09U3Y8	
DT	01-MAY-2000 (TrEMBLrel. 13, Created)	
DT	01-MAY-2000 (TrEMBLrel. 13, last sequence update)	
DT	01-MAY-2000 (TrEMBLrel. 13, last annotation update)	
DE	STARP ANTIGEN.	
GN	STARP.	
OS	Plasmodium falciparum.	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX	NCBI_TaxID=5833;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FCCL1/HN;	
RA	Shan Z.X., Yu X.B., Li X.R., Ma C.L.;	
RT	"Cloning and characterization of Plasmodium falciparum FCCL1/HN isolate	
RL	STARP gene".	
SR	Submitted (DBC-1999) to the EMBL/genbank/DBJ databases.	
DR	EMBL; AF209925; AAF21035.1; -	
SO	SEQUENCE 559 AA; 62167 MW; B5C1A99F0CA22C5D CRC64;	

Query Match	5.48;	Score 155.5;	DB 5;	Length 559;
Best Local Similarity	20.28;	Pred. No. 0.14;		
Matches 114;	Conservative 78;	Mismatches 242;	Indels 131;	Gaps 23

```

0Y 81 -----ASPSYDEKVGRIIDKANSPEARHPHLVAAPYIVHVDENII 1222
Db 98 YTNTRLPANNSTYTSYTKVTKNNKTNIKLIGJNNSTTINTSTENTSATKVKYENVTNQI 1577
0Y 123 LSKNEOSTQMTDESETRTISKNTSRTHTSEVGAENAEHANTSTSTHTSEVGAENAEVH 1828
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 158 LTGN NNTTNTTSTTEHHNNNTINTNTNSDNSNT--NINLNDNTSYTKKLNDIN----- 2088
0Y 183 AVADHSLAGERTMAET--MGLNTADTARLANIRYV--NTGAPRYNVLEPSTSLV 2377
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 209 ---TTONLTSTNTTSTVDNNNTNINFKRIDNNMTDIXSTWDYNTGKETQDNKMTDIKATD 2655
0Y 238 GRNQTATIKKKNOSILAPNNYFSKULAPALNAODFESTPTITMANYO---FLE 2933
Db 266 NNNNTTTTDTNNTNIVISYDNSKRTIVISYDNSKRTVTSYDNDMDTILTIDNDNMNMDILT 3255
0Y 294 LEKTQALDTDYVGNATATNFENGGRVDTGSMSEVPUQIOWTARIIPNCKDLNV 3533
Db 326 NNNDTDTISTDMDNA-DTKATDNNNNNTNATKATDNNNTKAIISFDNNNT----- 3707
0Y 354 ERRILAAVPSDPLEETKPDMLKALKIAEFGNEPNCNLQOGKDITFDFNEDQOTSQ 4133
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 371 --KTASTDNNNNTKAIISD-----NNNT--KTISTDNNNTKAIISTDN 4088
0Y 414 IKNOQLAELNATNTYVLDKIKLANKNMILIDRKFHYDRNNIAVGADESYYKAEHREVIN 4739

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:52:29 ; Search time 44.8 Seconds

(Without alignments) 456.674 Million cell updates/sec

Title: US-09-747-521-4_COPY_178_735

Sequence: 1 DNLAPELKKSSNRKRRS.....LYISNPYKVNYAVTKENT 558

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	2758	96.1	764	1	PAG_BACAN
2	211.5	7.4	192	1	YPA_BACAN
3	150	5.2	1230	1	SMC3_YEAST
4	141	4.9	1658	1	YMG7_YEAST
5	139	4.8	1358	1	SIR4_YEAST
6	138.5	4.8	1251	1	RBP2_PLAIV
7	138.5	4.8	1928	1	MYSL_YEAST
8	138	4.8	1385	1	FAT1_SCHPO
9	136	4.7	1790	1	USO1_YEAST
10	136	4.7	1803	1	YTL3_YEAST
11	134	4.7	753	1	YBV2_YEAST
12	134	4.7	1570	1	P3K1_DICDI
13	132.5	4.6	678	1	SIR1_YEAST
14	132	4.6	1208	1	PCP1_SCHPO
15	130.5	4.5	633	1	IPAA_SHIFL
16	130.5	4.5	803	1	SMIG_YEAST
17	130	4.5	2334	1	MAPA_BACSU
18	130	4.5	2710	1	TOXA_CLODI
19	129.5	4.5	719	1	YMA1_YEAST
20	129.5	4.5	2869	1	RBP1_PLAIV
21	129	4.5	1182	1	CGA2_HELPY
22	128.5	4.5	1271	1	Y338_MYGE
23	128.5	4.5	1460	1	N159_YEAST
24	128	4.5	976	1	SCP1_HUMAN
25	128	4.5	1167	1	CAGA_HELPY
26	127	4.4	1744	1	TANA_XENLA
27	126	4.4	1487	1	MDS3_YEAST
28	125.5	4.4	1630	1	MSPI_PLAFK
29	125.5	4.4	1639	1	MSPI_PLAFW
30	125	4.4	1726	1	MSPI_PLAFP
31	124.5	4.3	635	1	PIB2_YEAST
32	124.5	4.3	678	1	YNC7_YEAST
33	124	4.3	597	1	PTPX_CANAL

ALIGNMENTS

RESULT	1	STANDARD	PRT	764 AA.
PAG_BACAN				
ID	PAG_BACAN			
AC	P13423			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	PROTECTIVE ANTIGEN PRECURSOR (PA) [CONTAINS: PA-20; PA-63].			
OS	PAG.			
OG	Bacillus anthracis.			
OG	Plasmid pX01.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/staphylococcus group; Bacillus.			
ON	NCBI_Taxid=1392;			
ON	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=9132080; PubMed=1651334;			
RX	RA Singh Y., Kimpel K.R., Quinn C.P., Chaudhary V.R., Leppla S.H.,			
RA	Welkos S.L., Lowe J.R., Eden-McCutchan F., Vodkin M., Leppla S.H.,			
RA	Schmidt J.J.;			
RT	"Sequence and analysis of the DNA encoding protective antigen of			
RT	Bacillus anthracis.";			
RL	Gene 69:287-300(1988).			
RL	[2]			
RP	DOMAINS.			
RX	MEDLINE=9132080; PubMed=1651334;			
RA	Singh Y., Kimpel K.R., Quinn C.P., Chaudhary V.R., Leppla S.H.;			
RT	"The carboxyl-terminal end of protective antigen is required for			
RT	receptor binding and anthrax toxin activity.";			
RL	J. Biol. Chem. 265:15493-15497(1991).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).			
RX	MEDLINE=97192099; PubMed=9039918;			
RA	Petosa C., Collier R.J., Kimpel K.R., Leppla S.H., Liddington R.C.;			
RL	Nature 385:833-838(1997).			
RL	"Crystal structure of the anthrax toxin protective antigen.";			
CC	"FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,			
CC	AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE			
CC	DEATH. PA IS THOUGHT TO BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC			
CC	CELLS, THEREBY FACILITATING THE INTERNALIZATION OF LF OR EF. PA			
CC	ASSOCIATED WITH LF CAUSES DEATH WHEN INJECTED. PA ASSOCIATED WITH			
CC	PROTECTIVE ANTIGEN (PA) INDUCES IMMUNITY TO INFECTION WITH ANTHRAX.			
CC	SUBUNIT: ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT PROTEINS, A			
CC	PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN EDEMA FACTOR			
CC	(EF). NONE OF THESE IS TOXIC BY ITSELF. ONCE ACTIVATED, PA FORMS			
CC	HEPTAMERS WHICH INSERT INTO MEMBRANES AND FORM CATION-SELECTIVE			
CC	CHANNELS.			
CC	-1- SUBCELLULAR LOCATION: SECRETED.			
CC	-1- DOMAIN: THE C-TERMINAL PART OF PA IS REQUIRED FOR RECEPTOR BINDING			
CC	AND TOXIC ACTIVITY.			
CC	-1- PTM: PROTEOLYTIC ACTIVATION BY URIN CLEAVES THE PROTEIN INTO TWO			
CC	PARTS, PA-20 AND PA-63, THE LATER HEPTAMERIZE.			
CC	-1- SIMILARITY: TO C. PERRINGENS TOTA-B TOXIN AND TO VIP1 TOXINS IN			
CC	BACILLUS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M22589; AAA22637.1; -
DR PDB: IACC; 11-FEB-98.
DR InterPro: IPR003896; Binary_tox.
KW Toxin; Plasmid; Calcium-binding; Signal; 3D-structure.
FT SIGNAL 1 29
FT CHAIN 30 764 PROTECTIVE ANTIGEN.
FT CHAIN 30 196 PA-20.
FT CHAIN 197 764 PA-63.
FT DOMAIN 30 287 DOMAIN 1, CALCIUM-BINDING.
FT DOMAIN 288 516 DOMAIN 2, HEPTAMERIZATION.
FT DOMAIN 517 764 DOMAIN 3.
FT CA_BIND 206 206
FT CA_BIND 208 208
FT CA_BIND 210 210
FT CA_BIND 217 217
SQ SEQUENCE 764 AA; 85811 MW; 84BB22690FEAAB5 CRC64;

Query Match 96.1%; Score 2758; DB 1; Length 764;
Best Local Similarity 96.8%; Pred. No. 1,8e-142;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

OY 1 DNLQPELKQKSSNRKRSAGPTVPDRNDGIPDSLEVGYYVDVKNKRTFLSPWIS 60
DB 178 DNLQPELKQKSSNRKRSAGPTVPDRNDGIPDSLEVGYYVDVKNKRTFLSPWIS 237
OY 61 NHEKKGLTKYSSPEKNTASDPYSDPEKVTGRIDKNVSPARHPVLAATYIVHDMEN 120
DB 238 NHEKKGLTKYSSPEKNTASDPYSDPEKVTGRIDKNVSPARHPVLAATYIVHDMEN 297
OY 121 ILSKNEQDSTONTSETPTTISKNTSTSRHTSEVHGAHEVANSTSTHTSEVHGAHE 180
DB 298 ILSKNEQDSTONTSETPTTISKNTSTSRHTSEVHGAHEVANSTSTHTSEVHGAHE 357
OY 181 VHAVALDHSLSLAGERTVAETMGLTADTARLANIRVYNTGTAPLYNVLPPTSLVLGKN 240
DB 358 SSTVAIDHSLSLAGERTVAETMGLTADTARLANIRVYNTGTAPLYNVLPPTSLVLGKN 417
OY 241 QTLATITAKENQSLQILAPNNYPPSKNLAPILANMODDSSPTITMNTYNOFLEKTKQL 300
DB 418 QTLATITAKENQSLQILAPNNYPPSKNLAPILANMODDSSPTITMNTYNOFLEKTKQL 477
OY 301 RLDTQVYGNATYNTFENGVRVDTGSNMSEVLPQIQETTRARIIFNGKDLNVERRIAAV 360
DB 478 RLDTQVYGNATYNTFENGVRVDTGSNMSEVLPQIQETTRARIIFNGKDLNVERRIAAV 537
OY 361 NPSDPLETTKPMTLKEALKIAFGNEPENGNIQYOGKDTTEFDNFNDQTSQNIKNQLA 420
DB 538 NPSDPLETTKPMTLKEALKIAFGNEPENGNIQYOGKDTTEFDNFNDQTSQNIKNQLA 597
OY 421 LNATITVYVLKIKILNAKNLILIRDKRFHYDRNNIANGADESVVNEAHEVINSSTEGLL 480
DB 598 LNATITVYVLKIKILNAKNLILIRDKRFHYDRNNIANGADESVVNEAHEVINSSTEGLL 657
OY 481 LNIDDKIRKISGYIYEIEDTGLKEVINDRYDMLNISLRDQGTFTDFEKKYNNKPLPY 540
DB 658 LNIDDKIRKISGYIYEIEDTGLKEVINDRYDMLNISLRDQGTFTDFEKKYNNKPLPY 717
OY 541 ISNPNYKVNYAVTKENT 558
DB 718 ISNPNYKVNYAVTKENT 735

RESULT 2
YPA_BACAN STANDARD; PRT; 192 AA.
ID YPA_BACAN
AC P13422;

DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE HYPOTHETICAL 21.6 KDA PROTEIN IN PROTECTIVE ANTIGEN 5' REGION.
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1392;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89172073; PubMed=3148491;
RA Welkos S.L., Lowe J.R., Eden-McCutchan F., Vodkin M., Leppla S.H.,
RA Schmidt J.J.
RT "Sequence and analysis of the DNA encoding protective antigen of
RT Bacillus anthracis."
RL Gene 69:287-300(1988).
CC -I- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M22589; AAA22636.1; -
DR HSSP: P13423; IACC.
DR Hypothetical protein; Transmembrane.
KW TRANSMEM 162 182 POTENTIAL.
SQ SEQUENCE 192 AA; 21620 MW; DBC7150AE78F8AFA CRC64;

Query Match 7.4%; Score 211.5; DB 1; Length 192;
Best Local Similarity 34.8%; Pred. No. 2.7e-05;
Matches 46; Conservative 31; Mismatches 34; Indels 21; Gaps 4;

OY 433 MNILIRDKRFHYDRNNIANGADESVVNEAHEVINSSTEGLLNIDDKIRKISGYIYEI 498
DB 1 MNILIRDP-VHYDNGNITGYVDSYLKNAKYKOILMWSGCVSLNIDEDVQALSGYMLDI 59
OY 499 EDTE-----GLKEVINRYDMLNITSLRQDKFTIDPKKYNDKLPLYSNP 545
DB 60 KRPSSHLNTPYTTITLAGKDSGVGELYRLS-----DGAGFLDKPKDENKRSIV-DG 112
OY 546 YKVNYYAVTKEN 557
DB 113 DDVYVAVTKED 124

RESULT 3
SMC3_YEAST STANDARD; PRT; 1230 AA.
ID SMC3_YEAST
AC P47037;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CHROMOSOME SEGREGATION PROTEIN SMC3 (DA-BOX PROTEIN SMC3).
GN SMC3 OR YJL074C OR J1049.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W303;
RX MEDLINE=97474309; PubMed=9335333;
RA Michaelis C., Ciosk R., Nasmyth K.;
RT "Cohesins: chromosomal proteins that prevent premature separation of
RT sister chromatids."
RL Cell 91:35-45(1997).
RN [2]
RP SEQUENCE FROM N.A.

RA Rose M., Koetter P., Enlian K.D.:
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA Sor F.J.:
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
 CC PART OF A CHROMOSOME CONDENSATION MOTOR.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
 CC FLANKED BY PUTATIVE GLOBAL REGIONS AT THE N- AND C-TERMINUS.
 CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y14278; CAA74655.1; -
 DR EMBL: 249349; CAA89366.1; -
 DR EMBL: X88851; CAA61313.1; -
 DR SGD: S0003610; SMC3.
 DR InterPro: IPR003459; ABC_transport.
 DR InterPro: IPR003405; SMC_C.
 DR InterPro: IPR003395; SMC_N.
 DR Pfam: PF02483; SMC_C; 1.
 DR Pfam: PF02463; SMC_N; 1.
 DR Mitosis: ATP-binding; Coiled coil; Nuclear protein.
 FT NP_BIND 32 39 ATP (POTENTIAL).
 FT DOMAIN 172 482 COILED COIL (POTENTIAL).
 FT DOMAIN 685 1041 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1230 AA; 141336 MW; B152D88F780341F CRC64;

Query Match 5.2%; Score 150; DB 1; Length 1230;
 Best Local Similarity 20.4%; Pred. No. 0.73;
 Matches 18; Conservative 102; Mismatches 255; Indels 182; Gaps 28;

QY 3 LQPELKKOKSSNRKSTASGPTVPDRNDGIPDSLEVEGYTVYVKNKRTPLSPWISNI 62
 DB 371 LQPLASLQKRDILILKGEFARKSKDERDTWHSEI-----ELKSS-----IQNL 417
 QY 63 HE-KKGLTKYKSSPEKSTASDPYSDPEKYTGRIKKNVSEARHPLVAAYPIVHDMENI 121
 DB 418 NELESQLOMDRTSLRKQYSAID-----EEIEELIDISNGPDTGQL-----EDFDESLI 466
 QY 122 ILKKNEDQSTQNTDSETRTISKNTSTSRHTSEYHGNAEYHANTSTSRHTSEYHGNAEV 181
 DB 467 HLKQKLESELDTRKELMRKQKLOTLETLLSDVNON-QRNVNEMTSRS-----LANGII 520
 QY 182 HAAVIDHSLSLAGE-----RTNAETMG-----LNTADTARLANNIR 217
 DB 521 NKEITTEKLITSPESVFGTIGELIKYNDKRTCAEVLIGGNSLPHIVDTETATLNNEL 580
 QY 218 YVNTGAPIVNVLPSTSLVGLKNOTLATIRAKENQSLIAPNNVYPSKMLAPITALNAOD 277
 DB 581 YRMKGGGVTF--IPLNKLSLSDSVKFPSSNTTQIQFPLPKIKIYERFBEKA-----VKH 633
 QY 278 DESSTPTMYNOFLEKTKQLR--LDTDQV-YGNIAF--YVFENGRAVDTGSMWE 331
 DB 634 VFEKTIYVKNLGGGLKAKKHKLNATLTDGRADKRGVLGGLDQHKRTRLESILKMLNE 693
 QY 332 VLPOIOETARTIIFNGKDLNVERRIAANPS-----DPLETTKPD 372
 DB 694 SRGQHKILEELDFVRNELNDIDIKIQVNGNIRKVSNDRESVLTINIEVYRTSLNKKNE 753
 QY 373 -MTLKEALK-IAGFNPNGNLQYOGKDIETFPN---EDQOTSNIKQLA----- 419
 DB 754 KLLEESLNMLILKLEKLTNNRTFAQEKLTMTFENDLLQSFDSLSKKEKRLSELSTKEIS 813

QY 420 -----ELNATN-----ITYVLDKIKLAKNAILRDKRFHYDRNNIANG----- 458
 DB 814 AAHKKLNTSDALEGITTTTID--SLNALESKILIPENDLESKMSFVGADPFIQLODELK 871
 QY 459 -----ADESVYKAEARREVINST-----EGLL-----NIDK 485
 DB 872 ELQLEKSEVKEKHENAVLEIGTVOREIESLIAETNNKKLLERANNQORLLIKLNDFOK 931
 QY 486 DIRKILSGYIEIEDTEGLKE-----VINDRYDM-----IN-----ISSL 520
 DB 932 SVEKTMIKTKTLTVRRRELQORIREIGLPEDALVNFSDITSQDLQRLNDMNTFISGL 991
 QY 521 RQDCK-TFIDPKRYNDK 536
 DB 992 KNVKKRAFEFNKFKFNER 1008

RESULT 4
 ID YM67_YEAST STANDARD; PRT; 1658 AA.
 AC 003661; Q04988;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOThETICAL 187.1 KDA PROTEIN IN GUA1-ERG8 INTERGENIC REGION.
 GN YMR219W OR YMR261.13 OR YMR959.01.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID:4932;
 RN [1]
 RP SEQUENCE OF 1-711 FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
 RA Walsh S.V.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 608-1648 FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A.,
 RA Smith J., Johnston A., Whitehead P., Barrell B.G., Walsh S.V.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 249809; CAA89934.1; -
 DR EMBL: 249939; CAA90190.1; -
 DR SGD: S0004832; YMR219W.
 KW Hypothetical protein.
 SQ SEQUENCE 1658 AA; 187137 MW; 3893P968305A575D CRC64;

Query Match 4.9%; Score 141; DB 1; Length 1658;
 Best Local Similarity 20.4%; Pred. No. 3.4;
 Matches 125; Conservative 92; Mismatches 221; Indels 176; Gaps 31;

QY 11 KSSNSRRKSTASGPTVPDRNDGIPDSLEVEGYTVYVKNKRTPLSPWISNIH-E-K 65
 DB 488 KSEYSEVWLENETDPAIVERENQ-IND---VEGVDYGVKSVESDLHESP--DWLYDLAA 541
 QY 66 KGLTKYKSS-----PEKMTASDPY---SDPEKYTGR-IDKNVSEARHPLVAAYPIVHV 116
 DB 542 RAMLIQDQSRNSNCPQKEOVSEYSLGHNSGNSLGSGLDES---EEOIPLKD----- 591
 QY 117 DMENIILSKNEDQSTQNTDSETRTISKNTSTSRHTSEYHGNAEYHANTSTSRHTSEV 175
 DB 592 -----FTGNNNNNLATDRGLDLSVYEIEVEKYSKKLDOSTFKELVPLSTDTINNSS 644

```
QY 176 HGMAEVAHAVIDHSLAGERTWATMGLNTRA-----DTARLNANI 216
DB 645 LGNESTIYSLDDADAISENLTDVPLMEIKTTPKYEVIISSEVSSTSYEDNTVAMPPOV 704
QY 217 RYVNGTARIYVNLPTSTLVGKNQTLATIKAKENQLSQILAPNNYPSKNLAPIALNAQ 276
DB 705 EY-----TSPFNMD -PFNSL-----ND---DYKKHDLKSTLAA-----LAPAFYTKD 744
QY 277 DDFSSPTITM-----NYNOFLEKTKOLRLDTPDOVYGNIAFYNGFENGVRVDTGSN 328
DB 745 AEFVAGVTKSLTSTSGHTNIFHTSKETKQVS -DLDESTENVTPEENMTG----DENKN 799
QY 329 MSEVLPOQETTARIIFNGKDLNVERIAVANPSDPLETKPDTLKEALKIAGCFMFP 388
DB 800 QSKNFPVGNASTDKSTENTD---EKYFSALNTYN---VTGSSCEDDIETASNAVEE- 850
QY 389 NGNLQYQCKDITEFDFNF -DOOTSQ---IKNOLA-----ELNATNITYVLDKIKINA 437
DB 851 -NLATYCKDMNEMAMSSGDECVKQNDGSKTQISFSTDSPNPQESNDNTEFSSTK--- 905
QY 438 KWNILIRDKRPHYDRNNIAVGADESVEKFAHR-EVIN-----SST 476
DB 906 -----YKVRNSDLEDESLKELTKAEVVDKDEESEDSEYEQDYADEPGND 953
QY 477 EGLLNLIDKDIRKILSGYIVEIEDTEGLKEVINDRYDMLNISLQDQKTFIDFKKYNDK 536
DB 954 EGSNENIYKGTK-----DTLGIVEPENEKVN-----KVHEE 985
QY 537 LPLYISNPYKVN 550
DB 986 ETLFEANVSSSVNV 999

RESULT 5
SIR4_YEAST
ID SIR4_YEAST STANDARD: PRT: 1358 AA.
AC P11978:
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE REGULATORY PROTEIN SIR4 (SILENT INFORMATION REGULATOR 4).
GN SIR4 OR SIEP OR ASDI OR UTRH OR YDR227W OR YD9934.12.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
ON 111
RN 111
RP SEQUENCE FROM N.A.
RX MEDLINE=88142836; PubMed=3325825;
RA Marshall M., Mahoney D., Rose A., Hicks J.B., Broach J.R.;
RT "Functional domains of SIR4, a gene required for position effect
RT regulation in Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 7:4441-4452(1987).
RN 121
RP SEQUENCE FROM N.A.
RX MEDLINE=95192063; PubMed=7885847;
RA Davies C.J., Hutchinson C.A. III;
RT "Insertion site specificity of the transposon Tn3.";
RL Nucleic Acids Res. 23:507-514(1995).
RN 131
RP SEQUENCE FROM N.A.
RX STRAIN=5288C / AB972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A.;
RL Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: THE PROTEINS SIR1 THROUGH SIR4 CONTROL THE EXPRESSION OF
CC MATING TYPE GENES RESIDENT AT LOCI OF EITHER END OF CHROMOSOME
CC III. SIR3 AND SIR4 ASSOCIATE WITH THE C-TERMINUS OF RAPI TO FORM A
CC DNA-BINDING COMPLEX THAT INITIATES THE REPRESSION AT THE HM LOCI
CC AND TELOMERES.
CC -1- SUBUNIT: INTERACTS WITH RAPI C-TERMINUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M37249; AAA20881.1; -
DR EMBL; U13239; AAC33144.1; -
DR EMBL; 248612; CAAB8507.1; -
DR PIR; A29360; A29360.
DR SGD; S0002635; SIR4.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein;
KW Coiled coil.
FT DOMAIN 1277 1347 COILED COIL (POTENTIAL).
FT VARIANT 994 994 P->L.
SQ SEQUENCE 1358 AA; 152061 MW; 9C698765964F094E CRC64;

Query Match 4.8%; Score 139; DB 1; Length 1358;
Best Local Similarity 19.3%; Pred. No. 3.3;
Matches 121; Conservative 107; Mismatches 234; Indels 164; Gaps 24;

QY 2 NLOJPELQKSS-----NSRKRSTSGPTVPDRNDGIPDSLEVEGYVD 47
DB 24 NDKIFREKSEKNEVTKPIPLTFPAKSKNYGRSPAIHTSHQSSDVKPTS----- 74
QY 48 VKNKRFLSPWISNIHEKGLTKYKSSP-EKWSIASDPYSDFEKT----- 92
DB 75 -----HKLOQPKSSPLKNNVNSPFSNLKISNKLKLSLRKTS 117
QY 93 GRID-KNVSPEARHPLVAAYPIVHVDMENILSKREDSTONTQSE-TRTISKNTSRT 150
DB 118 GRIESNNSHDSASRLAS-----FEOTAFSRHAQOOTSTFNSKPVRTIVP-ISTSQ 168
QY 151 HTSEVHGNAEVAHANTSTRTHTSEVHGNAEVAHAVIDHSLAGERTWATMGLNATADTA 210
DB 169 NNSFLSGVKSLLSEKIRDSKELG-----INLANQVPLE----- 205
QY 211 RLMANIRYVNGTARIYVNLPTSTLVGKNQTLATIKAKENQLSQILAPNNYPSKNLAP 270
DB 206 -----KPLKGSADI-----GASVSLTKDKSIRKDYEEKKEKLNKFNASDSLSV 255
QY 271 IALNAQDFFSPTITM-----NYNOFLEKTKOLRLDTPDOV 308
DB 256 PKVSAGDSGISPEESKARSPGIARPNALQTEVYGINEESTNERLEINQEKPVKLDENSAN 315
QY 309 GNIAFYNGFENGVRVDTGSNNSSEVLPOQETTARIIFNGKDLNVERIAVANPSDPL-- 366
DB 316 STVASALDTNGTSATTELTLSKKIYP-----SPKVAIDQDKITTLHDEKTLAPSKHQPT 371
QY 367 -ETTKPDMTLK--EALKIAGFNEP-----NGNLQYQCKDITEFDFNFDOOTSNI 414
DB 372 EQKMKEDADLKRMEILKSPHLSPADRPQGRNRNRNSTDEETTKLAFLEVEGQEN- 430
QY 415 KNOQLAELNATNITYVLDKIKLNKAKNNILIRDKRPHYDRNNIAVGADESVEKFAHREVINS 474
DB 431 -----NYNSTSRST-----EKKNDMNTSAKN--NGENKKKIGKRPELMSSTEAHANVTE 478
QY 475 STEGLLNIDKDIRKILSGYIVEIEDTEGLKEVINDRYDMLNI--SSLRQDQKTFIDFKK 532
DB 479 ETKQIOQSVRIDGRKVL-----QKVG-ESHIDSRNNTLNVTPSKRPOLGEIPNPKK 530
QY 533 Y--NDKLPYISNPYKVNVAATKE 556
DB 531 HKPNEGRTPNISNGTINIQKLEPRE 556

RESULT 6
RBP2_PLAVB
ID RBP2_PLAVB STANDARD: PRT: 1251 AA.
AC 000799;
DT 01-APR-1993 (Rel. 25, Created)
```


DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE RETICULOCYTE BINDING PROTEIN 2 (FRAGMENT).
 GN RBP2.
 OS Plasmodium vivax (strain Belen).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN NCBI_TaxID=31273;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92315338; PubMed=1617731;
 RA Gallinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
 RT "A reticulocyte-binding protein complex of Plasmodium vivax
 merozoites.";
 RL Cell 69:1213-1226(1992).
 CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
 CC HUMAN RETICULOCYTE CELLS.
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND (PROBABLE).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M88098; AAA29744.1; -
 KM Malaria: Receptor; Membrane.
 FT NON_TER 1
 FT NON_TER 1251
 SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;

Query Match 4.8%; Score 138.5; DB 1; Length 1251;
 Best Local Similarity 21.8%; Pred. No. 3.1;
 Matches 127; Conservative 96; Mismatches 227; Indels 133; Gaps 31;

QY 7 ELKQSSNRKKRSTSGATVPDRD-----NGIPDSLEEGYT--VDVKNK----- 51
 DB 756 ELKFESEKNVNVLETENMSKNTNELDVHKNIDQYKVALELDLSDELTKQKSSKLIEM 815
 QY 52 --RTPLSPWISNIHEKKGLTKYKSSPE---KWTASDPYDFEKYTGRIKVNVSPEARH 105
 DB 816 GNGIYIAKVLLINQYKNK-ISSIKSKEAVSVKIGNVSKKHSELSKIT-----CSQKSYD 868
 QY 106 PLVAVAPYIVHDMENITLSKNEQSTQNTDSETRTISKNTSRTHTSEVHGAENVANT 165
 DB 869 NITIALEK--QTEQLNLRNSETQEKNTNSDSKLEIKITDFESLKNALKTLEG--EYNALK 924
 QY 166 STSRHTSEVHGAENVAVAHIDHSLSLAGEPTAEFTMGINTADTALNINIRVNGTAR 225
 DB 925 ASSDNH-EHVQSKSE---PVNPALSEI-EKEETDIDSLNTA---LDELKKGRCEVS 974
 QY 226 IYNNVL--PTTSVLGKNQTLATIKAKENQLSQILAPNNYPSKRLAPIALNAODDFSSTP 283
 DB 975 RKLIIDYTKELSDDELINTI---EKNVKAYLA---YIKKN-----YEDVYQDV 1019
 QY 284 ITMNTNQLELEKTKOLRLDQVGNINATYVNEGNRVRVGTGSNMSEVLPOIQETTARI 343
 DB 1020 LRLN-----EHNPTKVY-----SNHEPTNF-----DKSNKSEELTKAVTDSKTI 1059
 QY 344 INGGKOLNVERRIAVNPSDPLETKPRMTLKEALKIAFGFENRNGNIOYQCKDTEED 403
 DB 1060 ISKLKV-----ITVENTEMTENTESSAKETLEA-----TNE-----LKNKTSLSNET- 1103
 QY 404 FNFDOOTSINIKNOIALNA-----TNIY-TVLDRKIKLAKNANILIRDKRFYDNNIA 456
 DB 1104 ----YQTSNEVKLQEMKSNADKYIDVSKIFNTYLDTOQKSIYVN-----QHSINN- 1150
 QY 457 VQADESVVQEAHREYVNSSTEGLLINIDKIKILSGYVEIEDTGLEKIVINDRYD--- 513
 DB 1151 ----KDKLKGKLOELIDADDSFTLESI-KKFNEIYSHIKTINIGELBOLQOTNKSEHDNVA 1205

QY 514 -----MLN-ISSLRODGKTFIDFKKYNDKPLIYSPNPK 547
 DB 1206 KHKKEIVHLINRVESLKGDKVKNH-DDDQYKKKILNASLNDNIK 1247

RESULT 7
 ID MYSL_YEAST STANDARD; PRT: 1928 AA.
 AC P08964;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MYOSIN-1 ISOFORM (TYPE II MYOSIN).
 GN MYO1 OR YHR023W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae;
 OC NCBI_TaxID=4932;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=91088308; PubMed=2263482;
 RT Sweeney F.P., Watts F.Z., Pocklington M.J., Orr E.;
 RT "The MYO1 gene from Saccharomyces cerevisiae: its complete nucleotide
 RT sequence.";
 RL Nucleic Acids Res. 18:7147-7147(1990).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=94378003; PubMed=8091229;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
 RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
 RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
 RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
 RA Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
 RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
 RA Vaudin M.;
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT VIII.";
 RL Science 265:2077-2082(1994).
 RN
 RP SEQUENCE OF 1-760 FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=88111539; PubMed=3322809;
 RA Watts F.Z., Shields G., Orr E.;
 RT "The yeast MYO1 gene encoding a myosin-like protein required for cell
 RT division.";
 RL EMBO J. 6:3499-3505(1987).
 CC -!- FUNCTION: REQUIRED FOR CELL DIVISION.
 CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X53947; CAA37894.1; -
 DR EMBL: X06187; CAA29550.1; -
 DR EMBL: 010399; AAB68872.1; -
 DR PIR: S05806; S05806.
 DR PIR: S12323; S12323.
 DR PIR: S46773; S46773.
 DR HSSP: P08799; 1MND.
 DR SGD: S0001065; MYO1.
 DR InterPro: IPR000048; IQ.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00063; myosin_head; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR ProDom: PD000355; myosin_head; 1.

DR	SMART: SM00015; IO: 1.	
DR	SMART: SM00242; MYSC: 1.	
DR	PROSITE: PSS00066; IO: 1.	
KW	Myosin: Actin-binding; APP-binding; Coiled coil; Alkylation.	
FT	DOMAIN 1 793	MOVIN HEAD-LINE.
FT	DOMAIN 794 823	COILED COIL (POTENTIAL).
FT	DOMAIN 856 1911	APP (BY SIMILARITY).
FT	NP_BIND 180 187	ACTIN-BINDING (BY SIMILARITY).
FT	DOMAIN 460 529	ALKYLATION (BY SIMILARITY).
FT	MOD_RES 703 703	ALKYLATION (BY SIMILARITY).
FT	CONFLICT 36 36	K -> I (IN REF. 1).
FT	CONFLICT 46 46	I -> T (IN REF. 1 AND 3).
FT	CONFLICT 59 59	V -> S (IN REF. 1 AND 3).
FT	CONFLICT 86 86	L -> F (IN REF. 1).
FT	CONFLICT 330 330	MISSING (IN REF. 1 AND 3).
FT	CONFLICT 343 343	N -> S (IN REF. 1 AND 3).
FT	CONFLICT 421 425	COAKET -> PTLSSL (IN REF. 1).
FT	CONFLICT 515 515	D -> S (IN REF. 1).
FT	CONFLICT 529 535	SKGPPG -> ARGDR (IN REF. 1 AND 3).
FT	CONFLICT 541 541	D -> V (IN REF. 1 AND 3).
FT	CONFLICT 550 551	TD -> LM (IN REF. 1).
FT	CONFLICT 573 573	R -> A (IN REF. 1).
FT	CONFLICT 582 582	H -> D (IN REF. 1).
FT	CONFLICT 588 599	ETVGEWLSKNN -> NTLNKAQYPTK (IN REF. 1).
FT	CONFLICT 599 599	MISSING (IN REF. 3).
FT	CONFLICT 627 632	ESSSSA -> GNLNLC (IN REF. 1 AND 3).
FT	CONFLICT 695 695	R -> S (IN REF. 1 AND 3).
FT	CONFLICT 736 742	ENSTYTT -> RKNPHD (IN REF. 3).
FT	CONFLICT 756 756	E -> R (IN REF. 1 AND 3).
FT	CONFLICT 773 784	NKPKPKGVLA -> ILTVOKLEYS (IN REF. 1).
FT	CONFLICT 793 794	KL -> NV (IN REF. 1).
FT	CONFLICT 896 900	N -> T (IN REF. 1).
FT	CONFLICT 900 900	N -> NSQITKNTNITLTPSTYIGERPKRVICGN (IN REF. 1).
FT	CONFLICT 906 906	N -> I (IN REF. 1).
FT	CONFLICT 911 911	N -> Q (IN REF. 1).
FT	CONFLICT 915 930	NSGLNRAVTSSETIQ -> RIATKILKPAINIT (IN REF. 1).
FT	CONFLICT 934 939	DDLJSE -> MTLFL (IN REF. 1).
FT	CONFLICT 951 953	ADN -> RRI (IN REF. 1).
FT	CONFLICT 955 958	EAHH -> KKLD (IN REF. 1).
FT	CONFLICT 1002 1002	S -> C (IN REF. 1).
FT	CONFLICT 1049 1049	L -> D (IN REF. 1).
FT	CONFLICT 1056 1056	C -> S (IN REF. 1).
FT	CONFLICT 1060 1060	M -> I (IN REF. 1).
FT	CONFLICT 1085 1085	V -> E (IN REF. 1).
FT	CONFLICT 1123 1123	A -> C (IN REF. 1).
FT	CONFLICT 1133 1133	L -> S (IN REF. 1).
FT	CONFLICT 1144 1146	KSN -> NLI (IN REF. 1).
FT	CONFLICT 1159 1168	REKREDOKK -> TRKEEDDK (IN REF. 1).
FT	CONFLICT 1179 1181	SKT -> ELKV (IN REF. 1).
FT	CONFLICT 1184 1185	LE -> WK (IN REF. 1).
FT	CONFLICT 1188 1204	LSOETSLNLYNKRISG -> CHRYKLSILKOKNIR (IN REF. 1).
FT	CONFLICT 1224 1224	P -> S (IN REF. 1).
FT	CONFLICT 1228 1228	E -> Q (IN REF. 1).
FT	CONFLICT 1253 1253	E -> O (IN REF. 1).
FT	CONFLICT 1311 1323	POKESDINKLME -> LTKSLILTNGNAS (IN REF. 1).
FT	CONFLICT 1400 1400	D -> H (IN REF. 1).
FT	CONFLICT 1454 1554	SEDLRLQKLDSESTREKELSSITKQKQOFQENCMDDLOGG NEIRLREHHTHQAQAEVYNKMASTIEKLTONKOKKLLIM BEMRENSMDHLOETLLE -> P (IN REF. 1).
FT	CONFLICT 1566 1566	D -> V (IN REF. 1).
FT	CONFLICT 1630 1646	DLKQLODHTTKVEMLN -> SEARSILYSGNAV (IN REF. 1).
FT	CONFLICT 1698 1704	MISSING (IN REF. 1).
FT	CONFLICT 1725 1737	PTOLOMEOSNRG -> NTFANGTFKEM (IN REF. 1).
FT	CONFLICT 1754 1757	FODE -> LMM (IN REF. 1).
FT	CONFLICT 1777 1777	D -> E (IN REF. 1).
FT	CONFLICT 1788 1788	R -> T (IN REF. 1).
FT	CONFLICT 1825 1825	S -> D (IN REF. 1).

FT	CONFLICT	1882	1882	FWK -> NSGCRDADD; (TN REF. 1)
FT	CONFID	1902	1904	Supplied (JUN-1997) to the EMBL/GenBank/DBJ databases.
SO	SEQUENCE	1928	AA; 223634	MM; 6F5AC76114F3DC9F CRC64;
	Query Match	4.8%;	Score 138.5;	DB 1; Length 1928;
	Best Local Similarity	18.8%;	Pred. No. 5.7;	
	Matches 118;	Conservative 114;	Mismatches 219;	Indels 177; Gaps 28;
Qy	4	OLPELKQSSNSRRKRSRSGAPTVVDRNDGIPDSELEVBG-YTVVYKKNKPTFLSPWISNI	62	
Db	1097	ELKKKKEDIDHKKKELAFSS---KQD---DAVSEHGKITTALKEKTRI-----	1138	
Qy	63	HEKKKLTGKYSKSPKSWASTADPDYDFEKVYTGKIDKNSVPEARHPYAAVPIYHVMENII	122	
Db	1139	---QLTEKSNKYQK---IKKEYSNFORET---KEQDKKRNLSV-----ESLN	1177	
Qy	123	LSKNEQSTQNTDSEFTPTISKNTSFRHTSEVHGNAEYHANTSPKTHTS-----	173	
Db	1178	DSKIELEBAR-----LSQEISLNQYLNKRISGNS-VETNISTSTRSTSYSDPDLKE	1228	
Qy	174	-----EVHGNAEVHNAVAIDHSLSLAERPMWAE-MGLNTPADTRLANIRY	218	
Db	1229	DIKKYVYLQAFETITRNLENE---IEKKKLIRLRTETRIASSSFEDOKIRYAOMK	1285	
Qy	219	VNTGTAPIYNNVLPETSLVGLKQNOTLATIKAKENQLSQILAPNNYPSKRLAPIALNAODD	278	
Db	1286	LKKLIQDMDPSIPDSIIL---NEPLDNCPRKSDIKMLLEVDYL-KRQIDLETIRAHND	1340	
Qy	279	FSSPTITANNYNQLELEKTKQLRLDTDOYGNINATYNNENGKRVARDTGSNMSEVLPQIE	338	
Db	1341	-AENAIISALHSKFRKIQ---GESSLSDDIY---KLKFEASERKVSLEDKTKTPTPLDR	1393	
Qy	339	T---TARITFGKQDNLNERRIAAANPSPDLETPTKPMITLKEALKIAGFNEPNNGLOY-	394	
Db	1394	TNLPGVDIKNNDISIKKKEEIRYKK---LENYK---LQELI-----NESNKKLSQL	1439	
Qy	395	-----QCKDITPEPDNFDQTSQNIKNOL-----	418	
Db	1440	TLDLRQSKSEKALLSEQDRLOKDLSTEROKELLSSTIKQOKQOFENCMQDLOGNEURL	1499	
Qy	419	-----AELNATNITVLDKIKL-NAKNMIIILRDKRFHYDRNNINAVAGDESIVKKAH	468	
Db	1500	REHINALQAEDVDYKNNASIIIEKTKQKOKREKLIMEREM--ERRDSMQQLOETILE---	1554	
Qy	469	REVINSSTEGGLLNIDKIRKILSGIYVE-----IEDTEGLKEVINDRYDMLNISSLR	521	
Db	1555	-----LKRQVQYKTKILSDLAHLKRLSAVEDRSQYDEINRLKEELN-CSLK	1601	
Qy	522	QDGKTFIDEKKYNKDLPLYSINPNKKVN	549	
Db	1602	AETNLKKEPAFLKYLETSTNDSEAKIS	1629	
RESULT 8				
FAT1_SCHPO STANDARD: PRT: 1385 AA.				
ID	FAT1_SCHPO	STANDARD:	PRT: 1385	AA.
AC	013735; Q9UTJ0;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	ACTIN INTERACTING PROTEIN 3 HOMOLOG.			
GN	FAT1 OR SPAC15A10.15 OR SPAC15E1.01.			
OS	Schizosaccharomyces pombe (fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomycetes.			
OX	NCBI_TaxID=4896;			
BN	[1]			
RP	SEQUENCE OF 1-1033 FROM N.A.			
KC	STRAIN=972;			
RA	Murphy L., Harris D., Wood V., Bartell B.G., Rajadream M.A.;			
RL	Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.			

FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4B18 CRC64;

Query Match 4.7%; Score 136; DB 1; Length 1790;
Best Local Similarity 18.2%; Pred. No. 7; Mismatches 204; Indels 150; Gaps 25;
Matches 104; Conservative 114;

QY 2 NLQPELKOKSSNSRRKKRSTAGPTVPDRDNDGIPDSLE-VEGYTVDVK---NKRRTFLSP 57
DB 1229 NLQPELKOKSSNSRRKKRSTAGPTVPDRDNDGIPDSLE-VEGYTVDVK---NKRRTFLSP 1272
QY 58 WISNHEKKGLTKYKSSPEKWTASDPSPDFKYVGRIDKNVSPARHPLVAAYPIVAVD 117
DB 1273 EYSELED-----KLKASEDKNSKYLELOKESKEIKEEDLAKTE-----LKIQ 1315
QY 118 MENIT-LSKNEDOSTQNTDSETRTSKNTSTRTSTSEVHNAEVAHANTSTSTSTSEVH 176
DB 1316 LEKINTNLKAKKKS-----ESELRLKKTSSSEERK----- 1345
QY 177 GNAEVAHAAIDHSLSLAGEPRTWAEITMGINTADTARLANNIRVYNTGTAPIYVNLPTTSLV 236
DB 1346 -NAEQLEKLEKKEIQIKKQAFERKRLLEGGSTTIOEYSEKINT-----LEDELI 1395
QY 237 LGKNTLTATIRAKE-----NOLSOILAPNNYPSKNLAPIALNADDDSS--TPTTMNTN 289
DB 1396 RLQNN--ELAKKEIDNRSELEKYSLSNDELLEKONTIK-SLODELLSYKDKITTRNDE 1452
QY 290 QLELEKTKQLRLDQVYGNATYVNFENGVRVDTGNSMSEVLPOIOETPAR---IFPN 346
DB 1453 KILSTERONKRLDESLEK-----EOLRAOESKAKVEGGLKLEESSEKAELEK 1502
QY 347 GKDL-----NLVERIAAVNPSDPLET--TKPDMLKEALKIAFGNEPENGLOYGKDDT 400
DB 1503 SKEMAKKLESTIESNETELKS--METIRKSDKLEQSKSA---EEDIKNLQHEKSDLI 1557
QY 401 EEDFDFDOTSNIKNOI-----AELNAT--NIYVLDRKILKAKNNILIRDRFRHD 451
DB 1558 S-RINESKDELEELSKRIEAKSGSELETYKQELNNAOERIRIAEENTVLKSK----- 1611
QY 452 RNNIIVGADSEVYKAEHREINSSTEGLLNIDKIRKILSGYIEIED-----TEG 503
DB 1612 -----LEDIERELKAKKAE---IKNSQEKELTIRLRLLELDSDTOQKAK 1656
QY 504 LKEVINDRYDMINISLRQDKTPIDEPKKYND 535
DB 1657 SEERRAEVRKFOVEKSQLDEKAMLETKYND 1688

RESULT 10
YJL3_YEAST STANDARD; PRT; 1803 AA.
AC P47024; P87192;
DT 01-FEB-1996 (Rel. 33, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSPOSON T14 207.7 KDA HYPOTHETICAL PROTEIN.
GN TYAB OR YJL113W OR J0780.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97103775; PubMed=8948101;
RA Czaplewich C., Kordes E., Pujol A., Jauniaux J.-C.:
RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
RT reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,
RT SPI10, GCD14, REP1, PHO86, NCA3, ASF1, CCT7, GZF3, two tRNA genes,
RT three remnant delta elements and a Ty4 transposon.";
RL Yeast 12.1471-1474(1996).
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: 249389; CAAB9409.1; -
DR SGD: S0003649; YJL113W.
DR InterPro: IPR001584; Rye.
DR InterPro: IPR001878; ZnF_CCHC.
DR Pfam: PF00665; Iwe; 1.
DR SMART: SM00343; ZnF_C2HC; 1.
KW Transposable element; Hypothetical protein.
SQ SEQUENCE 1803 AA; 207693 MW; 16DCD728448D52D3 CRC64;

Query Match 4.7%; Score 136; DB 1; Length 1803;
Best Local Similarity 20.5%; Pred. No. 7.1;
Matches 138; Conservative 91; Mismatches 247; Indels 198; Gaps 33;

QY 1 DLQPELKOKSSNSRRKKRSTAGPTVPDRDNDGIPDSLEVEGYTVDVKNKRRTFLSPMIS 60
DB 1168 DGNNTLKEKLVYDKNVPPTDNGTVSPR-----LEQNTIASGSPQYQVYNSAPFLKERS 1219
QY 61 NIHERK-----GLTKY-----KSPKWTASDPYSDFEYVGRIDK 97
DB 1220 SLNMRKRKRHRDKNNSLSYELERDKKSKKNRVALLIDNMETVAPRIATYVEALSK 1279
QY 98 NVSPARHPLVAAYPIVAVDMENTILSKNEDOSTQNTDSETRTSK--TSTSRTHTSEV 155
DB 1280 NDELEKHEHYKQAY--IKELQNLKMKYFDVDVYRSSE--IPDNLVPTNTTFTYKR 1333
QY 156 HG--NAEVAHANTSTSTRTSTSEVHNAEVAHAAIDHSLSLAGEPRT--WAETMGINTADT--AR 211
DB 1334 NGTYKARIYCRDQTSPPDTSYTTTESLNHHIKITFLMIANNRMFMKTLDINHFLYAK 1393
QY 212 LNANIRVYNTGTAPIYVNLV--TTSVLGKNTLTATIRAKENO-----LSQI-LAPNN 261
DB 1394 LEE-----IYIHPHRCVVKLNKALYGLKQSPKRMNDHLQYLGILKONS 1443
QY 262 YIP-----SKNLAPIALNADDESSTPTIMYNOFL-ELEKTKDLR-----LQTD 305
DB 1444 YTPGLYQTEDEKMLM--IAVYVDCVIAASNEORLDEFINKLKSFLKITGTLIDVDLTD 1502
QY 306 QYVYGNATYVNFENGVR-----RVDTGNSMSEVLPOIOETPARITFNGKDLNLEVR 356
DB 1503 -ILGMDLYNKRRLGITIDLTAKSFINRMD--KYNELEKIRKRSSTPHSTYK----- 1551
QY 357 IAAVNP--SDPLETTKPD-----MTIKEALKIAFGNEPENGLOYGKODITEFDFNFDOOT 410
DB 1552 --IDPKKDVLOMSEEEERQGLKIQQL-----GELNYV-RHRCRYDIEFAVK 1597
QY 411 SGNITNOLAELNATYIVYVLDKIKLNAMNNILIRKRRHYDN-----NIANGAESVYK 465
DB 1598 VARLVNYHERFYIMYKIQYL-----VRYKDIGIHYDRCKDKKVIATDASVGS 1650
QY 466 EAHRE-----VI-----NSSTEGLLNIDKIRKILSGY----- 494
DB 1651 EYDAQSRIGVILIMYGMNIFNYISNKSSTRCVSSTAELHAITEGTADETLKVTLKEIGE 1710
QY 495 -----IVEIEDF-----EGLK-----EVINDRYDMINISLRQDKTFT-- 528
DB 1711 GDNNDIVITDSKPAIQGLNRSYQOPKREKFWIKREIIKEKIKESIKLKITGCGNIAD 1770
QY 529 -----DREKY 533
DB 1771 LITKPVASDFKRF 1784

RESULT 11
YBV2_YEAST

ID YBV2_YEAST STANDARD; PRT; 753 AA.
AC P38261;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOHETICAL 85.5 KDA PROTEIN IN VPS15-YMC2 INTERGENIC REGION.
GN YBR102C OR YBR0831.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=95208357; PubMed=7900426;
RA Mannhaupt G., Stucka R., Ehme S., Vetter I., Feldmann H.;
RT "Analysis of a 70 kb region on the right arm of yeast chromosome II.";
RL Yeast 10:11363-1381(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X78993; CA55605.1; -
DR EMBL; Z35971; CAAB5057.1; -
DR PIR; S44682; S44682.
DR SGD; S0000306; EX084.
KM Hypothetical protein.
SQ SEQUENCE 753 AA; 85537 MW; 1BCDE834466C3536 CRC64;

Query Match 4.7%; Score 134; DB 1; Length 753;
Best Local Similarity 19.9%; Pred. No. 2.8;
Matches 129; Conservative 93; Mismatches 227; Indels 198; Gaps 30;

QY 63 HEKGLT---KYSSPE---KSTASDPSDFE-----KVTGRIDKN 98
DB 15 HVKSSASPKOKTPPPAPKOKTKKPYSDLKDPATSYTLPIINARERSVATSMQRR 74
QY 99 VSPEARL--PLVAAYPIVHDMENIILSKNEDOSTONTDS-----E 137
DB 75 LSHNTNYAPPTLDYSMPLEDMPMIYPNDNVDSHNSSTTNESSYSGPNSLNL 134
QY 138 TRTISKNTS-----TSRTHSEVHGNAEVHANTSTSRTHSEVHGNAE--VH----- 182
DB 135 TADLSLNDSSYKNVPARSAMRNVTNPGSNDPFGNSTSLRKLMLANPHFNKAKDFHDKLGN 194
QY 183 --AAVADHSLSLAGEKRWAEKMGINTADTARLANINITYVTGTPATYVNLPTTSLVYGN 240
DB 195 ASAITTDKFTSNLJD-----LSIQVEEYKLNINKSYNEIMT--VNNDLVAMLEKLR- 245
QY 241 QTLATIKAKENOSQILAPNNYPSKMLA--PALNAODEEST-----PTM 286
DB 246 -----VAANINDLNEVDQCTKAIEKRLQDOIDEROGFENNVESHNSPALLPKA 300
QY 287 NYN-----QLELEKTKROLRDTDOYVYGN-----ATYVFNFGVRVDTGSMGEVL 333
DB 301 GONGNLMBRRDSSVLLIEKFMDEL--DQLEKFNVEGAQKFNSTKGHILMNSANWME-- 356
QY 334 PQOETIAR-----IIFNGDLNVERR-----IAAVNSDPLETTKPKDTLKEAL 379
DB 357 --LNTTGGKPLQWQVIFLNDLVLIADKSRKNDPFIVSOCYPLKDVTVVQEEFSTR-- 411
QY 380 KIAFGFEPNGNLQYQSKDITEFD-----FNPOQOSONKKNOLAELNA 423
DB 412 RLLEKFSNSNSSL-YECRADDECSRLLDVTRKAKODLCDFHVEEENSKRRESFRYLOS 470
QY 424 T-----NIYVLDRKIKLNAKNINILIRDKRFHYDRNNI 455

DB 471 TQQTGRENRRNSPNKRRKRRSGSITPGRNVTGAMDYLLQ-NLTLSMHSRPRSDMST 529
QY 456 AVG---ADESVK---FAHREVINSSTEGLLNLDKDKRLKLSGYVIEDETEGLKEVIN 509
DB 530 AQRKLFDEGVEEIDIELARLFESAVETLL-----DIESOL-----EDLSERIS 574
QY 510 DRYAM-LNITSLRODGKFTIDFKKYNKRLPLYISPNKVVAVYATK 555
DB 575 DEELMLNLISLKEQRR-----EAISSKLSOSIISNNEIVHLKSGTE 617
RESULT 12
P3K1_DICDI
ID P3K1_DICDI STANDARD; PRT; 1570 AA.
AC P54673;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PHOSPHATIDYLINOSITOL 3-KINASE 1 (EC 2.7.1.137) (PI3-KINASE)
DE (PBDINS-3-KINASE) (PI3K).
GN PIKA OR PIK1.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxId=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=96009592; PubMed=7565716;
RA Zhou K., Takegawa K., Emr S.D., Firtel R.A.;
RT "A phosphatidylinositol (PI) kinase gene family in Dictyostelium
RT discoideum: biological roles of putative mammalian p110 and yeast
RT Vps34p PI 3-kinase homologs during growth and development.";
RL Mol. Cell. Biol. 15:5645-5656(1995).
CC -1- CATALYTIC ACTIVITY: ATP + 1-PHOSPHATIDYL-ID-MYO-INOSITOL = ADP +
CC 1-PHOSPHATIDYL-ID-MYO-INOSITOL 3-PHOSPHATE.
CC -1- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U23476; AAA85721.1; -
DR DictyDb; DD01099; PIKA.
DR InterPro; IPR002420; PI3K_C2.
DR InterPro; IPR000341; PI3K_ras_bld.
DR InterPro; IPR001263; PI3Ka.
DR InterPro; IPR000403; PI3_P14_kinase.
DR Pfam; PF00613; PI3Ka; 1.
DR Pfam; PF00792; PI3K_C2; 1.
DR Pfam; PF00794; PI3K_rbd; 1.
DR Pfam; PF00454; PI3_P14_kinase; 1.
DR SMART; SM00145; PI3Ka; 1.
DR SMART; SM00146; PI3K; 1.
DR SMART; SM00144; PI3K_C2; 1.
DR SMART; SM00144; PI3K_rbd; 1.
DR PROSITE; PS00915; PI3_4_KINASE_1; 1.
DR PROSITE; PS00916; PI3_4_KINASE_2; 1.
DR PROSITE; PS50290; PI3_4_KINASE_3; 1.
KW Transferase; Kinase; Multigene family.
FT DOMAIN 41
FT DOMAIN 59
FT DOMAIN 73
FT DOMAIN 161
FT DOMAIN 168
FT DOMAIN 294
FT DOMAIN 304
FT DOMAIN 308
FT DOMAIN 324
FT DOMAIN 413
FT DOMAIN 424
FT DOMAIN 503
FT DOMAIN 510
FT DOMAIN 579
FT DOMAIN 821
FT DOMAIN 828
FT DOMAIN 831
FT DOMAIN 836
POLY-ASN.
POLY-ASN.
POLY-ASN.
POLY-ASN.
POLY-THR.
POLY-THR.
POLY-ASN.
POLY-SER.
POLY-ASN.
POLY-ASN.
POLY-ASN.
POLY-ASN.

FT DOMAIN 1309 1570 PI3K/PI4K
SQ SEQUENCE 1570 AA; 178374 MW; 558678872B34D783 CRC64;

Query Match 4.7%; Score 134; DB 1; Length 1570;
Best Local Similarity 21.4%; Pred. No. 7.5;
Matches 109; Conservative 68; Mismatches 205; Indels 128; Gaps 19;

CC 12 SSNSKKKSTAGPTVPDRDNGIPDSLEVEGYVDVKNK-----RTFLSPWISNIHEK 65
DB 176 SSSGSDSSASNOPIIRTRNREGSILNLKQGLVKEISQRFQTPPTASTYTRNNANISIK 235
CC 66 KGLTYKSSPEKWSASDPYDFEYVGTGRIDKNVSPFAHPLVAAVPIVHVMEINILSK 125
DB 236 DKISLLEKQERRKODSE-VQGRKRVYVLADSSNIQIYHPSVL-----IEKMSKLDLT 288
CC 126 NEDOSTQMTDSEPTISKNTSRHTSEVHGNAEVHANTSTST-----RTHTSEVHGNAE 180
DB 289 EKKPATTTTTTTTTTSTSTSTSTPTTTTTT-----TNTSTNDITTKPKTSPTKNNEE 341
CC 181 -----VHAVIDHSLSLAGERTMAETMGL----- 204
DB 342 RSQSPITTPKOPVEEIVKVSSTPKS-NNTSKTSSDTPTGKTTRKDKKKDKKSRDSGN 400
CC 205 -----NTADTARLNANIRVNTGTAPY-----NVLPSTSLVIGKN-OPLATIKA-----K 249
DB 401 LYIVANTNTNTSSNNNNNNNNNNNNNETIIRKGRVLVTPSSDLKKNIQIYFTIPIPNVAK 460
CC 250 ENQLSQILAPNNYPSKNL---API-----ALNAODFSSPTITMANYNOFLEKTKOL 300
DB 461 TNKPNQLISNTSQOFLKTLISNEIPIDCKINDINDTAFSDLSASASSSF--IRKSSQS 518
CC 301 RLDTQOVGNATYVPENGRVR-VDTGSWSEVVLPOIQETTRARIIFNGKDLNVERRIA 359
DB 519 LL-----NVQSLRYKAIKTSFNILFLMPN-----QSKTKILO 549
CC 360 VNPSPLETTKPDMLKEALKIAPGE---NEPNGNLQOGKDIPEFDPNQDSQNTK 415
DB 550 VKGSPTIE-----NLKERIISDYLFNNNSNNNNNNCKGADSYLLIDFNDPMERSVLV 603
CC 416 NQLAELNATNTYVLDKIKILNKKMNLIRD 445
DB 604 NKSDYILDKRAQGLIPKLKVIKSTILLSD 633

RESULT 13
SIRL_YEAST STANDARD: PRT; 678 AA.

AC P21691;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE REGULATORY PROTEIN SIRI (SILENT INFORMATION REGULATOR 1).
GN SIRI OR YKR101W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_TaxID=4932;
DB [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91172215; PubMed=2005909;
RA Stone E.M., Swanson M.J., Romeo A.M., Hicks J.B., Sternglanz R.;
RT "The SIRI gene of Saccharomyces cerevisiae and its role as an
RT extragenic suppressor of several mating-defective mutants";
RL Mol. Cell. Biol. 11:2253-2262(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Gallion L., Dujon B.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN REPRESSING TRANSCRIPTION OF THE SILENT
CC MATING-TYPE GENES, NAT1 AND ARD1, POSSIBLY VIA POSTTRANSLATIONAL
CC MODIFICATION OF OTHER SILENCING PROTEINS.
CC -1- PPM: IT IS POSSIBLE THAT N-TERMINAL ACETYLATION OF SIRI IS

CC REQUIRED FOR COMPLETE SILENCING.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DB EMBL; M38524; AAA5046.1; -
DB DR 228326; CAA82181.1; -
DB PIR; S14173; S14173.
DB SGD; S0001809; SIRI.
KW Transcription regulation; Repressor; Acetylation.
SQ SEQUENCE 678 AA; 79929 MW; 6540628CEFC2P3BC CRC64;

Query Match 4.6%; Score 132.5; DB 1; Length 678;
Best Local Similarity 20.8%; Pred. No. 2.9;
Matches 103; Conservative 78; Mismatches 195; Indels 119; Gaps 24;

CC 77 KWSTA--SDPYSDPEKVTGRI---DKNVSPEARHPLVAAVPIVH--VDMENIILSKNEQ 129
DB 142 KKGTYIIRHMLKDKRKPLAGVCEFSUDVNPDRREYLI-EINSHQYIIIEGFLYLNERR 200
CC 130 STQNTDSETRTSKNTSTSRHTSEVHGNAEVHANTSTSRHTSEVHGNAEVAVAIDHS 189
DB 201 LYRWMDNNLR--SQGLTKMAHLRKTYNPVSLDIYSL-----NSNVEYFVKDDL 248
CC 190 LSLAGERMAETMGINTADTALNANIRVNTGTAPI--YVLPSTSLVIGKNQIATITKA 248
DB 249 FOLLCKRVFVFKCKV-----MENGCKGKAPLVRVARTT-----TAKA 286
CC 249 KE--NQLSQILAPNNYPSKNLAPALNAODFSSPTITMANY---QFLELEKTK-OL 300
DB 287 THIAAIAINSTRPDSFKSKNNDRYIVNEKPIVENTISLSDSDIKKQFTAEVYKKKI 346
CC 301 RLDTQOVGNATYVPE--ENGVRVDTGSWSEVVLPOIOE---TARIIFNGKDLNLY 353
DB 347 SADISQIEVHTQFNSQKEKNRIVNKS--SEVLDQISKFPVSRVTLTLLMSAGODKNVI 404
CC 354 E-----RRIAAVNPSPLETTKPDMLKEALKIAPGEFNEPNGNLQOGKDIPEFDPND 407
DB 405 ELVEELARLEKI---CIETQSL---EELRDFTQAN-----PEMQASFD 444
CC 408 QOTSONIKNQLAELNATNTYVLDKIKILNKKMNLIRDRFRHYDRNNTA---VGADESIV 464
DB 445 KEYYSIEEYKTTLELIEDLLITLIK--QMEMMAAEKKEFTSEYVSPRLVADGFLI 502
CC 465 KEAHREVINSSTEGILLNIDKIRKILSGYIEIDTEGLKEVINDRYDMLNLSLRQDG 524
DB 503 DLAEKRPINPK-DPRLTLTLKDHORAM-----IDOMNLV----- 536
CC 525 KTFIDFKKYNKDLPL 539
DB 537 --WPDFKRYDPIPL 549

RESULT 14
PCPI_SCHPO STANDARD: PRT; 1208 AA.

AC Q92351;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SPINDLE POLE BODY PROTEIN PCPI.
GN PCPI OR SPAC69.06C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
CC Schizosaccharomycetales; Schizosaccharomycetaceae;
CC
NCBI_TaxID=4896;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Flory M.R., Morpheu M., Joseph J.D., Means A.R., Davis T.N.;
 RT "Pcp1p, a Spo11p-related calmodulin target at the centrosome of the
 RT fission yeast *Schizosaccharomyces pombe*."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Connor R.E.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: SPINDLE BODY COMPONENT THAT BINDS CALMODULIN.
 CC OVEREXPRESSION OF PCP1 CAUSES THE FORMATION OF SUPERNUMERARY SPB-
 CC LIKE STRUCTURES AND DISRUPTS BOTH MITOTIC SPINDLE ASSEMBLY AND
 CC CHROMOSOME SEGREGATION.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z81317; CAB03608.1; -
 DR EMBL: AF348506; AAK31344.1; -
 KM Coiled coil.
 FT DOMAIN 151 375 COILED COIL (POTENTIAL).
 FT DOMAIN 387 803 COILED COIL (POTENTIAL).
 FT DOMAIN 874 1091 COILED COIL (POTENTIAL).
 FT DOMAIN 1177 1204 COILED COIL (POTENTIAL).
 FT SEQUENCE 1208 AA; 140763 MW; 70264159ADDA42424 CRC64;

Query Match 4.68; Score 132; DB 1; Length 1208;
 Best local similarity 20.0%; Pred. No. 6.7;
 Matches 109; Conservative 95; Mismatches 232; Indels 108; Gaps 23;

QY 2 NLOV-----PELKQSSNRKRRKSTAGPYVPHDNDGIPPSLEVEGTVDYVKKRFELS 56
 DB 203 NLOLQKRTESLQKSEKKNFLEKY--DYLSKVN-DVEGOSNVKFTETIR---FLE 255
 QY 57 PWISNIHEKGLTFKYSKSPKSTASDPYSDFEKVGRIKDNVSPEARHPLVAAYPIYHV 116
 DB 256 NALEKVGREKDSLSTEMBEDSKNEKVEYETRIQLQNLDE-LSEE-----L 301
 QY 117 DMENIILSKNEDOSTQNTDSETRISKNTSRTHTSEVHGNAEVHANTSTSRHTSEVH 176
 DB 302 DVADQLLTERED-----EIAFL-KRQIEKENSSEAFENEE---NSSYHLDG-- 345
 QY 177 GNAEVHNAVADHSLSLAGERTWATMGLNPTADTARLANN-IRYVNTGAPRIYVNLPTTSL 235
 DB 346 -----DYALILQKCEDEFADRIQVLTADLEKEKEMQIMHSEASIGLDSMOVHTL 395
 QY 236 VLGNQTLATIKAKENQSLIAPNNYPSKNLAPIALNAODRSPPTWNYVOFLE 295
 DB 396 QEOHAKNEEIEFLHDOISR-NEEGNFEDIML-----QFRSLE 434
 QY 296 KTYQLRLDQVGNIAFYNEGNRVAVDGSNNSEV-----LPQIQTETARIITFN 346
 DB 435 E-----ERDLESKLTQLEDNNSRLMTSSLCNQIESLRTQNRREIDEKHNHLRLASK 488
 QY 347 GKDLNTERRAAVNPSPDLETTPDMTILKALKAIFGFPNPNGLQYOGDITPEPFNF 406
 DB 489 NSDALEETNRIQOEYKLEETLR--MKNSMDLNEIHDLRENEGTLTKLIDSTIKENDRL 546
 QY 407 DOQTSQNIKN--OLAEINATNITYVLDKIK-----LNAKNNIL-IRD---KRFHYDRNN 454
 DB 547 INLEQRIKSYEVAVSELNGT-IDEYRNKRLKDKRETYNEVNAFOYKDNDRRHHESINK 605
 QY 455 IAVGADSVYKAEHREYVNSTEGLLNTIDK---IRKISGYVEIETDEGLKEVINDR 511
 DB 606 LQDREKELTSLNLEKKNLVSSLRVAMLEKERESIKKYLSSNKKADLDNT--NLMEIILNDK 664

QY 512 YDML 515
 DB 665 ISVL 668

RESULT 15
 ID IPAA_SHIFL STANDARD; PRT; 633 AA.
 AC P18010;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE 70 KDA ANTIGEN.
 GN IPAA.
 OS Shigella flexneri.
 OG Plasmid 210 kb invasion pWR100.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Shigella.
 OX NCBI_TaxID=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-M90T / SEROTYPE 5;
 RX MEDLINE=90221912; PubMed=2183200;
 RA Venkatesan M.M., Buysse J.M.;
 RT "Nucleotide sequence of invasion plasmid antigen gene ipaa from
 RT Shigella flexneri 5";
 RL Nucleic Acids Res. 18:1648-1648(1990).
 RN [2]
 RP SEQUENCE OF 1-88 FROM N.A.
 RC STRAIN-M90T / SEROTYPE 5;
 RX MEDLINE=89057927; PubMed=3057506;
 RA Venkatesan M.M., Buysse J.M., Kopecko D.J.;
 RT "Characterization of invasion plasmid antigen genes (ipABCD) from
 RT Shigella flexneri";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9317-9321(1988).
 CC -1- FUNCTION: ASSOCIATED WITH THE ENTRY OF THE BACTERIA INTO COLONIC
 CC EPITHELIAL CELLS.
 CC CC
 CC -1- MISCELLANEOUS: SYNTHESIS OF THIS IMMUNOGEN IS RERESSED AT 30
 CC DEGREES CELSIUS AND RESTORED AT 37 DEGREES CELSIUS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X17628; CAA36624.1; -
 DR EMBL: J04117; AAA26525.1; -
 DR PIR: E31265; E31265.
 DR PIR: S12763; S12763.
 KW Antigen; Plasmid; Virulence.
 SQ SEQUENCE 633 AA; 70093 MW; 2F804F45355E4751 CRC64;

Query Match 4.58; Score 130.5; DB 1; Length 633;
 Best local similarity 21.3%; Pred. No. 3.4;
 Matches 106; Conservative 68; Mismatches 193; Indels 131; Gaps 21;

QY 46 VDVKNKRTF--LSPWISNIHEKGLTKKSSPEKSTASDPYSDFEKVTGRIDKNVSPDEA 103
 DB 210 VDAQKLAFLFDSLNRQNKTTIDSKGFGIKSLDNLTVAVFP-ELIRKVLNLDLEDIKDS- 267
 QY 104 RHPVLAAYPIVHVDMENTILSKNEDOSTQ-----NTDSETRTISKNTSRTHTSEVHG 157
 DB 268 -HPIDQGLPTPEEDMPCDGPPIPGANEKTSQPYIHINHNDRRTDNRFEDNRVYDNSVHE 326
 QY 158 NAEVHANTSTSRTH-TSEVHGNAEVH-----AVAIHSLSLAGERTWATMGLNTA 207
 DB 327 NPENDAOPTQTDNLLSRNGNSLNLNFORALVQKVTYVLPSPHSIS----- 370

QY 181 YVEIGKILSRILSKINOPYKFLDVLTNTIKNASDSGODLLFTNQLKEHPTDSVEFLE 240
 Db 181 YVEIGKILSRILSKINOPYKFLDVLTNTIKNASDSGODLLFTNQLKEHPTDSVEFLE 240
 QY 241 QNSNEVOEFKAFAYEYEPQHRDVLQYAPAFNYMKFNEQELNLSLEELKQDRLMSR 300
 Db 241 QNSNEVOEFKAFAYEYEPQHRDVLQYAPAFNYMKFNEQELNLSLEELKQDRLMSR 300
 QY 301 YEKNEKIKOHQWMSDSLSSEGRGLKLLKQPIEPKKDDIHSLSQEEKLLKRIODSS 360
 Db 301 YEKNEKIKOHQWMSDSLSSEGRGLKLLKQPIEPKKDDIHSLSQEEKLLKRIODSS 360
 QY 361 DFLSTEEKEFLKLLQIDIRDSLSEBEKELNRIQVDSNPLSEKEFLKLLKIDIDPYD 420
 Db 361 DFLSTEEKEFLKLLQIDIRDSLSEBEKELNRIQVDSNPLSEKEFLKLLKIDIDPYD 420
 QY 421 INORLODTGGLIDSPSINLDVRKQYKRRDIONIDALLHQSISTLYLNKLYLENNINLT 480
 Db 421 INORLODTGGLIDSPSINLDVRKQYKRRDIONIDALLHQSISTLYLNKLYLENNINLT 480
 QY 481 ATTAGADVDSTDNTRKINGEJNEFKKNEKYSISSNMYIVDINERPALDNERLKMRIOLSP 540
 Db 481 ATTAGADVDSTDNTRKINGEJNEFKKNEKYSISSNMYIVDINERPALDNERLKMRIOLSP 540
 QY 541 DTRAGTYLNGKLLIQRNIGLEIKDVOIIOSEKEYIRIDAKVVPKSIDTKIOBAQLNIN 600
 Db 541 DTRAGTYLNGKLLIQRNIGLEIKDVOIIOSEKEYIRIDAKVVPKSIDTKIOBAQLNIN 600
 QY 601 QENKALGLPPTYKLTIFPNVHNRYASNVESAYLILNEMKNNISODLKKATNTLVDSNG 660
 Db 601 QENKALGLPPTYKLTIFPNVHNRYASNVESAYLILNEMKNNISODLKKATNTLVDSNG 660
 QY 661 RFVETDITLPLNIAQYHODEIYEQVHSGGLYVESRSILHGPSKGYELANDSEGFHE 720
 Db 661 RFVETDITLPLNIAQYHODEIYEQVHSGGLYVESRSILHGPSKGYELANDSEGFHE 720
 QY 721 FGNAVVDYAGYLLDKNOSDLVTSKKEFIDJFEKESNLTSGRTNEAEFAEFLMHST 780
 Db 721 FGNAVVDYAGYLLDKNOSDLVTSKKEFIDJFEKESNLTSGRTNEAEFAEFLMHST 780
 QY 781 DHAERLVKONAKPTFOFINDOIKETIINS 809
 Db 781 DHAERLVKONAKPTFOFINDOIKETIINS 809
 RESULT 2
 B59106
 hypothetical protein pXOI-122 - Bacillus anthracis virulence plasmid pXOI
 C:Species: Bacillus anthracis
 C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 09-Jun-2000
 C:Accession: B59106
 R:Okinkka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Kelm, P.; Koehler
 J. Bacteriol. 181, 6509-6515, 1999
 A:Title: Sequence and organization of pXOI, the large Bacillus anthracis plasmid harbori
 A:Reference number: A59091, M0ID:99445483
 A:Accession: B59106
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-800 <OKI>
 A:Cross-references: GB:AF065404; NID:94894216; PIDN:AAD3426.1; PID:94894338
 A:Experimental source: strain Sterne
 A:Note: similar to calmodulin sensitive adenylate cyclase, edema factor, cya, plasmid pX
 C:Genetics:
 A:Gene: pXOI-122
 A:Genome: plasmid
 C:Superfamily: calmodulin-sensitive adenylate cyclase; calmodulin-sensitive adenylate cy
 F:34-286/Domain: lethal factor amino-terminal homology <LFA>

Query Match 12.2% Score 504; DB 2; Length 800;
 Best local similarity 24.1% Pred. No. 3.3e-13;
 Matches 214; Conservative 154; Mismatches 297; Indels 224; Gaps 40;

QY 4 KKEFI-----KVISMCLVATITSGPVFIPLYQAGAGHGVDGMHYKEKKNKDKRDE 59
 Db 3 RNFETPNKFSIISVYLLFAISSQAIENVA-----MNEVYTSDIKRNHKT 49
 QY 60 ERNKTOEHLAEIMKHIVIEVGEAAVKKEAEKLEKVPDVLNEMAKGKIYIVDG 119
 Db 50 EKNKTEKFKFDSJNNLVKTEFTNETLTKIQOTDOLKKIKFDVLEIYSELGELYFEDI 109
 QY 120 DITKILSEALSEDKKKIKDIDYGDALLHEHYVAAKEGEYEVLYQSEEDYVENTAKLN 179
 Db 110 DLVEHKELODISEEKNMNSNGEKKVPASRPFVEKKRETKLII-NIKDAVINSQESKE 168
 QY 180 VYVEIGKILSRILSKINOPYKFLD--VLNTIKNAS--DSDGODLLFTNOLKE--HPT 232
 Db 169 VYVEIGKILSRILSKINOPYKFLD--VLNTIKNAS--DSDGODLLFTNOLKE--HPT 232
 QY 223 DPSVEFLQNSNEVOEFKAFAYEYEPQHRDVLQYAPAFNYMKFNEQELNLSLEEL 292
 Db 224 SIDINFIKENLTERQHAFLAFSYFADHRTVLELYAPDMFEYMNKL----- 271
 QY 293 KDQRLMSYER--MEKIKOHQWMSDSLSSEGRGLKLLKQPIEPKKDDIHSLSQEEK 350
 Db 272 -----ERKGEFKI-----SESLKEG-----VERDRIDL-----KGEKA 301
 QY 351 LKRIQIDSDPLSTEEKEFLKLLQIDIRDSLSEBEKELNRIQVDSNPLSEKEFLK 410
 Db 302 L-----KAGLYPEHADAKFT-----ARLNTYLFPRPNKLATN-----LI 339
 QY 411 KKLIDIDPYDINORLODTGGLIDSPSINLDVRKQY-----KRDIONIDAL-HQ-SIG 461
 Db 340 KSGVATKGLANVHGKSDPGVAGYIPFDQDSKHHQGLAVEKGNLEKKKSTHEGEGIG 399
 QY 462 STLKNTYL-YENNININLTITLQADVDSTDNTRKINGEJNEFKKNEK-----YSISSN 515
 Db 400 -----KIPKLKHRIEELK-----ENGILLKGEKEIDNGKRYLLESN 438
 QY 516 YMIVDINERPALDNERLKMRIOLSPDTRAGYLE-NGKLIQRNIGLEIKDVOI-IIOSEK 573
 Db 439 NQVYEF--RIDENNEVOYK-----TKEGKIYLVGEFNMNRNIEVAKNVEGLKPLTA 490
 QY 574 EYIRIDAKVVPKSIDTKIOBAQLNINOEMKALGLPKYTK-----LITFNVHNRYA 625
 Db 491 DY---DLFALAPSLTEIKQIPQ---KEMDKVNTPNSLEKQGYTNLLIKYGIERK-- 541
 QY 626 SNIVESAVLLNEMKNNISODLKKATNTLVDSNGRPVFDITLPLNIAEQYTHOD--EY 683
 Db 542 ---PDSTKGTLSNMOKOMDLRL-----NEAVKYYGYTGGDVNNHGTEDONEFP 587
 QY 684 EOVHSGGLYVESRSILHGS--PSKGYELANDSEGFHEFGHGVDDYA---GYL----- 732
 Db 588 EKQDEITLINEGEFLLTKNEMTGREFLEKNITGKDLIYFNRSYKNKLAEGNKAYIEWTD 647
 QY 733 -LDKNSDLVTSKKEF-----DIFKEGSGN-----LTSYGR 764
 Db 648 PITKAKINTIPTSAEFITKNLSIRSSNVGYYKSGDKDEPAKESYKTIAGYLSDDYNS 707
 QY 765 NEAEFFAE-----APRLMSTDAER-LKVQKNAKP--TFQFINDOI 803
 Db 708 ANHIFSOEKKRKKISIFRGIQAYNEIENVLKSKQIAPYKNYFOYLERI 756
 RESULT 3
 JS0029
 adenylate cyclase (EC 4.6.1.1) precursor, calmodulin-sensitive - Bacillus anthracis
 N:Alternate names: anthrax toxin edema factor
 C:Species: Bacillus anthracis
 C:Date: 31-Mar-1992 #sequence_revision 11-Nov-1994 #text_change 19-Jan-2001
 C:Accession: JS0029; PS0307; JS0602
 R:Robertson, D.L.; Tippetts, M.T.; Leppia, S.H.
 Gene 73, 363-371, 1988
 A:Title: Nucleotide sequence of the Bacillus anthracis edema factor
 A:Reference number: JS0029; M0ID:89211974
 A:Accession: JS0029

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:46:05 ; Search time 72.22 Seconds
(without alignments)
853.298 Million cell updates/sec

Title: US-09-747-521-2

Perfect score: 4145
Sequence: 1 MNKKKEIKIVISMCLVTAT.....KNAPKTFQINDIKFIINS 809

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR68:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4145	100.0	809	1 J00032	anthrax toxin lethal
2	504	12.2	800	1 B59106	hypothetical prote
3	497	12.0	800	1 J50029	adenylate cyclase
4	264.5	6.4	1979	2 C71622	hypothetical prote
5	263.5	6.4	2401	2 T28676	rhodopy protein -
6	255	5.9	2269	2 T28677	repeat organellar
7	246.5	5.9	1127	2 T18372	ORF MSV156 hypothe
8	245.5	5.8	1127	2 T28317	reticulocyte-bind
9	241	5.8	2829	2 A42771	protein g377 - mal
10	239.5	5.8	3119	2 T18414	centromere protein
11	235.5	5.7	2663	1 S28261	SCP1 protein - rat
12	234	5.6	946	2 S28061	synaptonemal compl
13	230	5.5	993	2 S49461	hypothetical prote
14	229	5.5	3724	2 T18427	hypothetical prote
15	224.5	5.4	1558	2 B71603	RESA-H3 antigen pr
16	222.5	5.4	2166	2 G70163	hypothetical prote
17	218.5	5.3	1163	2 D64315	type I restriction
18	218	5.3	756	2 C64236	protein V (fcrv) h
19	218	5.3	980	2 E71606	hypothetical prote
20	217	5.2	1088	2 T18559	hypothetical prote
21	216.5	5.2	1679	2 S48385	hypothetical prote
22	216	5.2	1005	2 A64465	hypothetical prote
23	215	5.2	2245	2 T18278	myosin heavy chain
24	214.5	5.2	1313	2 F96673	hypothetical prote
25	214	5.2	3225	2 I52300	giantin - human
26	214	5.2	3259	1 A56539	giantin - human
27	213	5.1	1516	2 E71619	RAD2 endonuclease
28	212.5	5.1	1957	2 T38077	hypothetical colle
29	211.5	5.1	1875	2 S38173	myosin-like protel

30	211	5.1	1252	2 B42771	reticulocyte-bind
31	210	5.1	1939	2 I48175	myosin heavy chain
32	209.5	5.1	1128	2 G86266	hypothetical prote
33	209.5	5.1	1780	2 T17272	hypothetical prote
34	209	5.0	1642	2 T08880	NMDA receptor-bind
35	208.5	5.0	691	2 S48390	hypothetical prote
36	208.5	5.0	841	2 A86188	hypothetical prote
37	208	5.0	1230	2 S56850	SMC1 protein homol
38	208	5.0	2710	2 A37052	toxins A - Clostrid
39	207	5.0	1922	2 T00637	hypothetical prote
40	207	5.0	3394	2 T18501	hypothetical prote
41	206.5	5.0	1650	2 T18444	hypothetical prote
42	206.5	5.0	1837	2 T41023	probable nuclear p
43	206	5.0	2139	2 T18296	myosin heavy chain
44	205.5	5.0	1738	2 T14867	interaptin - slime
45	204	4.9	1312	1 BMYDYL	RAD50 protein - ye

ALIGNMENTS

RESULT 1

J00032 anthrax toxin lethal factor PX01-107 precursor - Bacillus anthracis virulence plasmid

C.Species: Bacillus anthracis

C.Date: 31-Mar-1990 #sequence_revision 11-Nov-1994 #text_change 11-May-2000

C.Accession: J00032; C59104

B.Brady, T.S.; Robertson, D.L.

Gene 81, 45-54, 1989

A.Title: Nucleotide sequence and analysis of the lethal factor gene (lef) from Bacill

A.Reference number: J00032; MUID:90034185

A.Accession: J00032

A.Molecule type: DNA

A.Residues: 1-809 <BNA>

A.Cross-references: GB:M29081; NID:G143143; PIDN:AAAT9216.1; PID:G143144

R.Okinaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koeh

B. Bacteriol. 181, 6509-6515, 1999

A.Title: Sequence and organization of PX01, the large Bacillus anthracis plasmid harb

A.Reference number: A59091; MUID:99445483

A.Accession: C59104

A.Molecule type: DNA

A.Residues: 1-809 <OKT>

A.Cross-references: GB:AF065404; NID:94894216; PIDN:AAD32411.1; PID:94894323

A.Experimental source: strain Sterne

A.Note: Similar to Anthrax toxin lethal factor precursor; lef, plasmid PX01, B. anthr

C.Comment: This lethal factor of Bacillus anthracis is part of the tripartite protein

mer they cause anthrax, an infectious and often fatal disease of cattle, sheep, and o

C.Genetics:

A.Gene: lef; PX01-107

A.Genome: plasmid

C:Superfamily: anthrax toxin lethal factor; lethal factor amino-terminal homology

C:Keywords: toxin

F:1-33/Domain: signal sequence #status predicted <SIS>

F:34-809/Product: anthrax toxin lethal factor #status predicted <MAT>

F:44-295/Domain: lethal factor amino-terminal homology <LFA>

Query Match 100.0% Score 4145; DB 1; Length 809;
Best Local Similarity 100.0%; Pred. No. 3; le-158;
Matches 809; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MNKKKEIKIVISMCLVTATLTS	GPVFTPLVQAGGHDGMHVKKEKKDKKKRDEE	60
Db	1	MNKKKEIKIVISMCLVTATLTS	GPVFTPLVQAGGHDGMHVKKEKKDKKKRDEE	60
Qy	61	RNKTQEEHLKEIMKHVYKLE	VGEAVKKAERLLEKVPDYLEMKATGKTYIVDGD	120
Db	61	RNKTQEEHLKEIMKHVYKLE	VGEAVKKAERLLEKVPDYLEMKATGKTYIVDGD	120
Qy	121	ITRHISLEALSEDEKKIKTD	IGYGDALLHEHYVAKGGEYFVLYIOSSSEVDVEMTERALNV	180
Db	121	ITRHISLEALSEDEKKIKTD	IGYGDALLHEHYVAKGGEYFVLYIOSSSEVDVEMTERALNV	180


```

Db      DEEVCDDLKRRLSKESEMKMKKEEHDKRLAEKLDCCDVR--IRENNNEKNEDKINMLKEE- 737
Qy      460 IGSTLYKNKIYLENNANNINLTATGLGADLVSDTDNKK-IRGJIFNEKKKNFKXSIS- 514
Db      738 -----YED-KIN-----TLKEQUNEDKINTLKEQUNEDKINTLKEEYEHKINTMKEE 781
Qy      515 -NYMIVDINERPALD-----NERLKMRIOQSPDTRAGYL-----ENOK 551
Db      782 YEKKINTLNEQNEHKINTLNEQNEHKINTMKEEYEDKMTLNEQNEDKNSLKEEYENKI 841
Qy      552 LILQNNIGLEIDV--QIIOKESEKEYIRIDAKVVPKSIDTFKIOEQALINQ----- 601
Db      842 NQINNNNEKKIDVYNEYEEYDKLKVTLDEK---KKQYDKELINTYAHITAEHEQOJLLTE 898
Qy      602 -EMNKALGLPKYTKLITFENVNHRVANSIVESAYLILN-EMKNKIOS-DLIKVTYNTLVGD 658
Db      899 MEELKQCNQNRKYSDL-----YEKYI-KLIKSIQMIINIICQDIEHEDILIRREEYINN 952
Qy      659 NGRFVFTDITLPLNIAEQYTHODEIYEQVHSGLYVPESRSILLHGPSKVEYLNDSGFT 718
Db      953 KGL-----KKEVEKEHKR-----HSSFNLISKEKF- 979
Qy      719 HEFGHAVDDYAGYLLDKNOSDLVYNSKKF-----IDIFKEGSLN 758
Db      980 --FKRSIEBKSHELKKKHEKDLKSDKELEEKNNKIKRELINDIKKLQDEILVYKQ-SNA 1036
Qy      759 TSYGRTEAEFFAEAFALRMHSTPDHAEKRLTVKQKNAPKPTFOINDIQIKFI 806
Db      1037 QOYVHKKKSSWILLKQSKSEKIKDKRENOJINVEKNEEKDLKKKQDEINIL 1084

```

```

RESULT      5
            T28676
            rhoptry protein - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28676; A45521
R:Shha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A>Title: Comparison of two members of a multigene family coding for high-molecular mass
A:Reference number: Z20507; MUID:97077455
A:Accession: T28676
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: EMBL:U36927; NID:G1041784; PID:G1041785; PIDN:AA841263.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A>Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple c
A:Reference number: A45521; MUID:91101660
A:Accession: A45521
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KEE>
A:Cross-references: GB:M34281

```

```

Query Match      6.4% Score 263.5 Db 2 Length 2401;
Best Local Similarity 20.3%; Pred No. 0.0051;
Matches 193; Conservative 168; Mismatches 358; Indels 233; Gaps 42;

QY   2 NIKKEFIKIVISNCLVATITLSGPFVIPVOAGAGHDVMHMKREKKDKENRRDEER 61
     |||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db   1389 NIYEEFIYSYDLITHYLETVESKEPTIEYOIKKRIRTAQMELLINIKNVAKSKAYDDIEA 1448
     |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY   62 NKTOE-EHLKTIKMHI-----VKTEVGGE-----AVKKAELKLEKVPDVL 104
     |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db   1449 NEFDRIWTHFKKNLADVNDKFNEYSKVN-KGPDNISINSINNKKSTDENLLINIINQTK 1507
     |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY   105 EMYKRAIGGIYI-----VDG-DITHTIS---LEALSEPK 134
     |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db   1508 EMYANIVSKKIYSYVEAEENITINIPKLANSLNIOIKSSGIDLFENNINAILPYLDSDOK 1567

```

```

01 135 KIKIYKODLLEHHYVAAKEGEYVVL-VLOSSEDVYENKRLNYYEGLITLSDIL 193
02 1568 KDTLFIISPEKSTETTYIKIDSINTLIDLIRKSOLOAKKQOOLNLF-----NLLH 1622
03 194 SKINOPYKFLVDLNTIKNADSODQLLFTNOLKEHPTDSVEFLPONSNOVEOYPAKA 253
04 1623 DKV-OATNELKDTLSLDKKKEQ-----ILNKV-----LLHKSNELNKLCSNS 1666
05 254 FAY--YIEPQHDVLOLYAPAFNYMDKFNQOEIINLSLELKDORMLSRYEKMKIKOHY 311
06 1667 QNYDTILLESSKYDKRK--EKSNNNEKEKENLGINFDVKAKEO-FNNIDIKIEKLENNY 1722
07 312 OHMSD--SLSEEGKGL--KLO-----IPREP---KDIITHSOEEK-- 350
08 1723 KHSEKDNVNFSEENNIIOSKKLKLKELTNAFNAELKIEDKIERKNGILINKLETBKDCM 1782
09 351 -----LLKRIODSSD-----LSTEEKEFLKQIODISLSEEEKL-----INRI 393
10 1783 LETYKTLVELIKITKTDTKFTSATKFSKFLKYIDA-TSNSLINDOINTLOQTYVDLNOI 1841
11 394 Q-----VDSNPLSEKEKFLKKLKDIDOPYDINOQLDTGGLTDSPSINLOVRK 443
12 1842 NKHVASMAVADATNDNNNLIEKEKEATKTINMLTELE-----TIDSNKIDADGLH 1890
13 444 QYKRIQIMIDLHOSIS--TLYNKIYLENNM-----NMLTATLADL 487
14 1891 NKKIOLIYFNSLHSDISIKOLYKKHMAFKLNLGHINKKFKEDISKEFDNILOLSEEL 1950
15 488 VDSNTNTKINGIENEFKKNFKYSISSNYMVDINERPALNERLKWRIQSPDTRAGYL 547
16 1951 TANLNDLKEIOGKISDKKQOFLHALSET-PIPNFTLKEIYHDIYKKNQOIDELENTNE 2009
17 548 ENKGIILORNTGLEI-----KVOIIOSEKEYIRIDAKVVPKSIDTKI 592
18 2010 ENENITLIMDIITKLMKRVOSILNEVTTYENDSNIKOHODNNENNEDVSKI-KESLETTI 2068
19 593 OEAOLINOENMKALGKRYKTLTFENVNRYASIVYSAVLINEMKNNTIOSLIRKVT 652
20 2069 OSFOQILN-----KLEIKAKQIOPDNINNINNVISTISODVNDVKKHLISKUL--TIE 2116
21 653 NYLVGNGRFVFTDITLPIAEOY--THODEIYEOVHSGKLYVESRSILHAGSKGVEL 710
22 2117 NELLIO-----IQKSLDKIKSTYDINSQDTKVKVNPITHDYVBOQKKIQQNNPNK---- 2165
23 711 RNDSEGFITHEGHAVDYAGYLLDKN-OSDL---VTNSK-----KFTIDIKREE 754
24 2166 -----DEIDDLIOEIVANNKSELEKPLTIINKKNQVTPILSRIDKVINLIKSE 2213
25 755 GSN--LTSYGRITNEAEFFAEAFRLMHSTDHAERL--KVOCNAPKKTQFIND 801
26 2214 YNNNDNVSYNAKKLE--EDANNITRDLDTSHNMJLNDLIQN---FKIJD 2259

```

RESULT 6
128677
Thopytry protein - Plasmodium yoelii
Species: Plasmodium yoelii
Date: 15-Oct-1999 #sequence-revision 15-Oct-1999 #text-change 01-Dec-2000
Accession: Y28677; C45521
R. Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
Title: A gene coding for a high molecular mass thopytry protein of Plasmodium yoelii
Accession: Y28677
Status: preliminary; translated from GB/EMBL/DBJ
Molecule type: DNA
Residues: 1-2269 <KEE>
Cross-references: EMBL:U28677; NID:Q457145; PID:Q457146; PID:AAA21304.1
R. Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
Title: Identification of the gene for a Plasmodium yoelii thopytry protein. Multiple
Reference number: A45521; MUID:91101660


```
Db 1443 GDDINCEKNDQAEISTLKDEIKKISMVGEEL-----NRKNSYDEKVK-----NL 1489
QY 559 GLEIKDVOI-----IKOSEK-----EYIRD 579
Db 1490 TNELKELKIRKKGEEMAEIAELNKLKNIKERNKSVKQNDSESSNNITTKDGKTEPEYVND 1549
QY 580 AKVVP---KSKIDTYIOEAQNLINOEWMKALGLPYTKLITFVNNRNASVINESAYILIN 637
Db 1550 DKICKDMKANLYLKEKEP-----DLMDNINSLER-----ENFRVMSIVKE----- 1590
QY 638 EWMKNIOSDLKRYTNLYVDNGR-----FVFTDI-TLPNIAEDTYHODEIYEBOVH 687
Db 1591 --AKNVQNDKIVGIYIYFKCKCEKELKNDMLVLCVLDLISLFLPNDNFVNLFEKIDKL 1648
QY 688 SKGLVVP-ESRSILHGPSKVELRN---DSEGEI---HEFGHAV-----D 726
Db 1649 WKQVYIPEIRILFLRFSPFLDKLRNYKCVNEEYVNNEREYSWALEQYTLFETASNLK 1708
QY 727 DYAGYLLDKNOSDLVYNSK-----KFIDI-----FKEEGSNLT 759
Db 1709 EMIVYLEKEKEDSCENNSSNFDRPKITDILNFSKDSIRLKTIAQLRKELFEREAKNIL 1768
QY 760 SYGRTNEAEFFAEFRMLHSTDHAEKLYOKNAKPTQF-INDQIKFTI 807
Db 1769 NYDQIILNKYHECLR-----KLKIVKNMARELDFTNYSSKPSI 1808

RESULT 8
T28317
ORF MSY156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T28317
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: Z20484; MUID:9102612
A:Accession: T28317
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1127 <AFO>
A:Cross-references: EMBL:AF063866; NID:94049647; PIDN:AAC97677.1; PID:94049717
C:Genetics:
A>Note: MSY156

Query Match 5.9%; Score 245.5; DB 2; Length 1127;
Best Local Similarity 20.6%; Pred. No. 0.01;
Matches 167; Conservative 126; Mismatches 263; Indels 253; Gaps 37;

QY 60 ERNKTQEE---HLKEIMKHIVKIEVKGSEAVKKEAEKLEKVPDYLEMYKAIGKTY- 115
Db 24 ENNKVSLDIINSLLEILNLT-----KFSKTIINELIKNNKIKVEKIFY 66
QY 116 -----IVDGDITKHSLEALSDEKKIKD-----YKGKDALLHEHYAAEGYEPV 161
Db 67 MHNQKINDYNIILQVLEIYENNEINCKICKEFPCKNPLY-----NITYKKKLYIYD 118
QY 162 LVIOSSDYENFEKALNVYE-----IGKILSRILSKINPY-----QKFLDVL 207
Db 119 L-----DYEEKKDELVINIEOKNAVDKI--NDIKNNVNNIHSNDETITTGKSTLIDIL 170
QY 208 NTKINASDSGDGLLFTNOLKEHPTDSVEFLQONSNEVOEFAKAFAYVIEPQHRDVLQ 267
Db 171 NKLKLIVSDEKOLI--EQIYKINNNKEIEF--KNIDNVQKEIK----- 211
QY 268 LVAPAEAFYWDKNEQENINLSLEELKDQRMLSRYEKWEKIKOHQHSDSLSEGRGLK 327
Db 212 -----KQDELNKLDE-----SKKEFIK----- 229
QY 328 KLOPIPEKKDDIHSIQEKKELKRIQIDSSPFLSTEEKEFLKLOIDRSLSEBK 387
```

```
Db 230 -----KQELNKTIDKQOEELIKKLNDEKEINENIDOKLLDQINSKI--NTLNENIK 280
QY 388 ELNRIQVDSNPPLSEKEKEFLKLLKLDIOPDINORLODGGJLIDSPSINLDVRKQYKR 447
Db 281 GVMN-LYTEETNNKISNQNELNKL-DSTIKSLDEKOLLD-----ELDK 322
QY 448 DIQNIDALLHOSIGSTLYNKIYLYENNMINNLGTATLGADLVSDTNKINRGIFENERKN 507
Db 323 NINNI-----TSLYNK-----SNFKITNIQOLLESSTLD-FNNANIN--INELKSK 365
QY 508 FKYSISSNYMIVDINERALDNERLKNRIQLSPPTRAGIYENGKLLIQRNGLKIDVOI 567
Db 366 IK-----LFD-NDQKLNNDITDQNNKIT-----DEFNNSRIKREKIDTEYKKITD 411
QY 568 IKOSEKEYIRIDAVVPESKIDTKIOEAQLINOEWMKALGLPYTKLITFVNNR----- 623
Db 412 IKNNNLQKLEESYK-----KIDBOTEYKKNKINNEYNDIYL-KNNNLQKLEENKKIDE 465
QY 624 --YASNIVESAYLILNEWK--NNTQ-----SDLKRYNYLYVDNGRFFVFTITLP 670
Db 466 QTEYKKNKINKEYNDITELKNNNLQKLEENKNINDKLTCLKNDIESNTLEF-----NKL 520
QY 671 NIAEQYIHODEI-----YEOVHSKGLVVPESRSLHGPSKVELRNDSCGFIFHGHA 724
Db 521 NISDFKDSREIAKLNTYEYEQI-----RKDLLENIKTNELKLSLSSLEQL 570
QY 725 VDDYAGYL-LDKNOSDLVYNSKFFIDFKEEGSNLTSYGRTN--EAFFAEFRMLHST 780
Db 571 YDSKKNILDGIDKITVNSLKEKNDKIDEF-----SNIEKFDLYNYIENKFIGNLSIINKI 626
QY 781 DHAERLKYOKNAKPTQFINDQIKFTINS 809
Db 627 IN-----NDQFEYKINS 638

RESULT 9
A42771
reticulocyte-binding protein 1 - Plasmodium vivax
C:Species: Plasmodium vivax
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 28-Apr-1995
C:Accession: A42771
R:Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
Cell 69, 1213-1226, 1992
A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
A:Reference number: A42771; MUID:92315338
A:Accession: A42771
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2829 <GAL>
A:Experimental source: Belem strain, merozoites
A>Note: sequence extracted from NCBI backbone (NCBIN:108114, NCBIP:108115)

Query Match 5.8%; Score 241; DB 2; Length 2829;
Best Local Similarity 18.9%; Pred. No. 0.05;
Matches 194; Conservative 161; Mismatches 326; Indels 346; Gaps 46;

QY 39 DVGNHVKKEKKNKQENR-----KDEBRNKTQEBHKEIMKHIVKI----- 79
Db 706 DINALIEVEKFTYENKSESTLEMLKDEE---MEKIQDAKETFAKLNFVSDKLTLYYT 761
QY 80 ---EVKGEAVKKEAEKLEKVPDYLEMYKAIGKI-----YIVDGD-ITKHI 125
Db 762 KMSAEVYNAEGIKKEIKELNQKQFENVHKKMKESDASFIRFEALQNSMOQYNOEGALIEKH 821
QY 126 SLEALSEDK---KKIDYIKDALLHEHYAAEGV---EPVLVIOSSDYENFTEK 176
Db 822 QNRSEKEEYEFKNESVEEDLSRETEQEYTKHKNNSSRKGEISAEITMAREVINIKES 881
QY 177 ALANYVEI-----GKLSIDILSKINQYQKRLDVLNTIKNA--- 213
Db 882 QLNHYGVIEKFFSLIGDQNEVSTAKALKEKIVSDSLDKRIDQYETEKEKTSVAVENTVST 941
```

214 --SDSGDGLL-----FTNQLKEHPTDFS----- 235
Db 942 IQLSLAIDSLRLNLSINCKKRYNTDIDLLSKITLREBQVKEMPRKGDGCGENTTAL 1001
Qy 236 -VEFLQNSNEVOEFAKFAFYIEPOHRDVLQLYAPEAFNWDKNEOEINLSLELKD 294
Db 1002 LKSLRLDKMKKINELNDRLNSLDTRKEDLLKFYSE-----SKSKIHLSKQDKPR 1052
Qy 295 QMLSLSEKWEKIKO-----HYQWMSDSLSEGGKLLKLIQI-----PIEPKKDIIIS 343
Db 1053 QPPLNIDIDEMEDIKRDVDELNVNQ-----VISENKVTLFKNNSVYIEAMHSHINTVAHG 1108
Qy 344 LSOEKEKELKRIQIDSSDELSTEKEEFLKKL-----OID-IRDSLSEEEKELLR--- 392
Db 1109 ITSNNKEILSKYKEVEDKLNLYQONEDYKKVKNPENKOLEARSGMS-KLKEVINKHS 1167
Qy 393 --IQVDS-SNPLSEK-----EFLKKLKIQIOPYDINRIQD 427
Db 1168 EMTQLESTANTLKSNAKGENEHDLNKTGQMDDIYEKLLKIAEELKEGTVNE-LKD 1226
Qy 428 T-----GGLIDSPSINDL-----VRQYKRDQNT-----DAL 455
Db 1227 ANEKANKVEPERNITIGHLEITVEKDKAGKVEEMNSLTKRIEKLQETSDSDSONEL 1286
Qy 456 LHSIGSTLYN-KIY-----LYENNMINLTATLGA 485
Db 1287 VTTSTITKLEMAKGYEDVTKRNEEDSOLREKAKSLETIDEMKLLVOQVNMNLSAQG- 1345
Qy 486 DLVSDTNTKIRNGI--FNEFKKPKYSISSNMYMIVDINERPALDNERLKW-RIQLSPD 541
Db 1346 -----NAGISKEINELKGVIELISTFYSSI-----LEYKKNSSSES 1382
Qy 542 TRAGYLENGKLL-----QNNIGLEIKDVQIIOK--SEKEYIRIDAKVVPKSID----- 589
Db 1383 VFFSOQLANEFTKAECEEKNASARLAEAELKQIYKLDYSDIDKVV--KKIEIGIKRE 1439
Qy 590 -TKIOEAOININQEMNKALGLPKYTKLITFNVN-----RYASNIYESAVLINEMK 640
Db 1440 ILKMKRSALT--WEES---EKFKOMCSSHMAKEGGKKIETYLKNNGDG-----K 1486
Qy 641 NNISDLIKKVTNYLVNGRPFVETDITL-----PNTAEQYTHODEIYEQVHSGLYV 693
Db 1487 ANITDSQMEVGNVYSKAEHAFHTVEAQVDKTAFCESIVAYVTKMDNLFNBSLMEVKV 1546
Qy 694 PESRSLILGPGSKVGLRNDSEGFHEFGHAVDDYAGYL-----LD 734
Db 1547 K-----CEKKNDK--AEKYSAKLKPYPGRIKARYSENERKISLEKAKAYE 1590
Qy 735 KNOS-----DLVTNSKKFIDIFKEE-GSNLTSYGR--NEAEFFAEAFRLMSTDHAEK 787
Db 1591 KRESSQLNDVSTKSLQIQINCRQQLDSVLSNIGRVKONALQYFDSADSKMSKSVLPISLG 1650
Qy 788 VOKNAPK 794
Db 1651 AEKSLDK 1657

RESULT 10
T18414
protein g377 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18414
R:Handman, E.; Osborn, A.H.; Symons, F.; Van Driel, R.; Cappai, R.
Mol. Biochem. Parasitol. 74, 143-156, 1995
A:Title: The Leishmania promastigote surface antigen 2 complex is differentially expressed
A:Reference number: 218933; MUID:96360472
A:Accession: T18414
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3119 <HAN>
A:Cross-references: EMBL:L04161; NID:g309687; PID:g309688; PIDN:AAC37257.1

Query Match 5.8%; Score 239.5; DB 2; Length 3119;
Best Local Similarity 19.9%; Pred. No. 0.065;
Matches 182; Conservative 153; Mismatches 313; Indels 265; Gaps 46;

Qy 27 FLPVQAGAGHD-----VGMHVKEKKNKDEKRRDEKRNKTOEHLKEIKHAKIE 80
Db 964 FTISGEENHNRRLKRIEALNLEKEMKKRFNEQOEORRRKKADE--DEMNETIQKHD 1021
Qy 81 VKGEFAVKKEAAKLEKVPDVLNWKALGKITYVDC-----DTRKHLSLEASLD--K 134
Db 1022 METSKLEKKEEDVYQODEEFD-----KQYAIQQEELFLRTRDSEGESDVPK 1071
Qy 135 KRIK-----DIYKDALHEHYVAKGYPVLIQSSSEDY-ENTKALNAYET- 184
Db 1072 DKVPPDGRSPDSFYNNATISSFHEKMEELYN--TSISSLYVVEINRKPDDYKELK 1129
Qy 185 -----GKILSRDILSKINQ-----PYQ-----KFLDVNTIKNASDSOG 218
Db 1130 SKTYPRFDDLTSGTKNCKNKLFOKLNMTIKDKREYQKNISQYKKKVIDIIDDIO--KKANG 1187
Qy 219 QDLFTNQLKEHPTDSVEFLQNSNEVOEFAKFA----- 254
Db 1188 KYIIQNLIEKIDIKGVDVRLSDRKFYKFKRVLKKRKMMLDPRFQKGAIRFIKD 1247
Qy 255 --AVYIEPOHRVLO-LYAPEAFNWDKNEOEINLSLELKOQMLSR---YEKMEKIK 308
Db 1248 LTTTMEEEYTKVLEDIYMEKKRYKEEYS-----KKRRLISSLDLEVANKQIK 1296
Qy 309 QHYQHSDSLSEEGRGKLLKQIPIEPKKDDIHSLSQEEKELKRIQIDSSDPLSTEER 368
Db 1297 EHY-HKVDITSEH-----KFOEIROMHRKIENTIHELTKEMVYQIIDLTVY----- 1344
Qy 369 EPLKLIQIDRSLSEEEKELNRIQVDSNP-----LSEKKEFLKKLIQIDOPYDIN-- 422
Db 1345 -----HOLENISHSELQALQONKNIPRLNVLLEKELITRKKKN-KP-DISTS 1391
Qy 423 -----ORLQDTGGLIDSPSINDVRQYKRDQIONIDALHOSIGSTLYNKIYLENMN- 475
Db 1392 SHATDEQVSDT--LIRGAHNGDITIGEDND-----EVLLIEDIQL--KTKMGDNQNO 1442
Qy 476 -----INNLTATLGAADVDS-----TDNTKIRNGIFNEKKKNKYSISSNMYMIVDINER 524
Db 1443 VGSILEKLNLS-----DQYQLQDKLVNVEDIYKLN-RNPKHYIEKHLHESKINRE 1493
Qy 525 PALDNERLKWRIQLSPTTRAGYLENGKLLIQNNIGLEIKVO-----IIKOS 571
Db 1494 KFI-----TKVDVLSNYSSTLEYWVKFLHDFQEMSEKDELKHLVELEE 1539
Qy 572 EKEYIRIDAKVVPKSIDTFOIEAO--LNIQENMKALGLPKYTKLITFNVNRYASNI 628
Db 1540 RKKYITLQIOT--RDTLSTIQNGEGDHINNNNNN-----VNRNNLKKOY 1583
Qy 629 VESAYVILNEMKNNI--OSDLIKRVT-----NYLVNGNRFVETDITLPNTAEQYTHODEIY 683
Db 1584 LKDDLEISKLGHILEVDIKKNIALEQINVLNN-----TNETYVDYRDLMPAPRIY 1637
Qy 684 ---EYVHSGKLYVPESRSLILHGPGSKVGLRNDSEGFHEFGHAVDDYAGYLLDKNSDL 740
Db 1638 PVSEDIYDITVWRDNTAV-----INNTLRHFRVMPDQIKYDYDHLI-----P 1681
Qy 741 VTNSSKKRFDIFKE-----EGSNLTSYGRTN-----AEFAEAFRLMSTDHAEK 791
Db 1682 VYVIKEL--TYKENLADKINANNIYEMNKFYHLHEEFF--YILKH---YVELRKIQDL 1733
Qy 792 APKTFQFINDQIK 804
Db 1734 GEVAIPSFEEENK 1746

RESULT 11
S28261
centromere protein E - human

Best Local Similarity 19.1%, Pred. No. 0.21;
Matches 180; Conservative 190; Mismatches 352; Indels 218; Gaps 42;

QY 45 KEKEKNKDEKRRKDEERKTOEHLKEIMKHIVIE-----VKGE 84
Db 497 KELOEKDEKKNKDKTIVNNNEEQDLDLRNINKEIESINNNDNNNNNNKKEFKKITE 556
QY 85 EAVKKAALKLEKVP---DYLEMYKA-----IGKIYVGDITK- 123
Db 557 HILNKESISKHKKSPSRDKKEIKKLYTNKEDSTFEELKELEITNNKVVNVEEIDIGS 616
QY 124 -----HISLEALSED-----KKIKDIYGDALLHEHYVYAKEGEVPLVIOSSDY 170
Db 617 NEDDEYIHLKENLKEDEANENYNDKENNKKTEILKSKNIEENKRILEEKLKGKNNI 676
QY 171 VENTKALNYYEIGKILSRDIL---SKINQ-----PYOKFLDY---LMTIKNASDSD 217
Db 677 FKDKER-----YNSLGEVILNEIQUINEENKINDIODGNISKOKIIQSSSRNTDFTIKDIS 732
QY 218 GODLFTQOLKEHPDSEVEFLQNSNEVOEYFAKAFAYIEPQHRDYLQYAPAFNYM 277
Db 733 LNDLEKERKRRKQHFIDNLVKADKNEISENINKICDNNINNIYDESINNIYDESINNI 792
QY 278 DKFNQOEINL---SLEELKDORMLSRYEKMEKIKQHQHWSDSLSEEGRLKIKQIPLE 334
Db 793 ---YDESINNIYDESINNIYDENINNIYD---ENINNIYDENINNIYDEGINIKICDNI-LE 847
QY 335 PKK---DDII---HSLQOEKELKRIQDISD-----FLSTEKEFLKQ 375
Db 848 NKNIKTNDIYQVEENNSIEKELMISLKNKDINTYMKFKNVDIFINKIKRESLAKID 907
QY 376 IDIRSLSEEEKELNRIQV-----SSNPLS-EKEKFLKRLKIDIOPIYDIN 422
Db 908 KNKDNNNDDEYIMDNFENDEFTINHKMEITNKELDPLEINTQNEFINLEIDIKKKKRYND 967
QY 423 QRLQDTGLISPS---IMDYRKQYKRDIONDALLHOSIGSTLVKITYL----- 471
Db 968 HFNNDADMFEYEMKILKKDKKKNKEQEPFTD---ETFGSLQSHKIKTKYKNGEKKHDK 1023
QY 472 ---ENMNI---NMLTATLGADVDSTNTKINRGIFNEFKKNFYSSISSNYMIVDINER 524
Db 1024 NNEEKNILYDENGQYVSVLSD-----HKIEDIOD-----LHSIQTN--ICDENNI 1068
QY 525 PALDNERKMKHIQSPD---TAGYIENGKLLQIRNIGLEIDVOYITKSEKYEIRIDAKV 582
Db 1069 EOINENSKKQVIGSTMEKNKDNEMKDNEMKDNEMKDNEMKDNEMKDNEMKDNEMKDN 1127
QY 583 VPKSKIDRKIQEADLINQENWKA-----LGLPKYTKLITFNV--HNRYASNI 628
Db 1128 DMEKKNDMEKKNMENENEMKSDIENEMKNEMKNEMKNEMKNEMKNEMKNEMKNEMKNEM 1187
QY 629 VESAYILINEMKN--IQSDLIK--VTNYLVGNGRFEV-----TDITLPIAE-QYTH 678
Db 1188 IENNMENNMENNMENNMENNMENNMENNMENNMENNMENNMENNMENNMENNMENNMEN 1247
QY 679 QDEIYQVHSGGLVPEPERSILLHGPSKGVELRD-----SEGFIHEFGHANDVAGY 731
Db 1248 QNNIQNNKH---VKDKNDLIN---NVDIINDVLKSDSKFEMMINISKEINIDFK-- 1296
QY 732 LLDKNOGLVNTN-----SKKFIIDFK---EGSGLTISYGTNEAEFFAEAFR 775
Db 1297 CVDKK-NDICNLNLELAKQSTNSNTTEKCIDHDFDYENKQVHDKIHEDREKYCESKL 1355
QY 776 LMHSTDHAER---LKVQKNAPK---TFQFINQIKFII 807
Db 1356 KYLPADDIKKMRSFIKIKKSKRENFLLISYMEETIYFLI 1395

RESULT 15
B71603
RESA-H3 antigen PFB0915w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000

C:Accession: B71603
R:Gardner, M.J.; Tetteelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.;
Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: B71603
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1558 <GAR>
A:Cross-references: GB:AE001424; GB:AE001362; NID:39845307; PIDN:AC71972.1; PID:9384
C:Genetics:
A:Gene: PFB0915w

Query Match 5.4%; Score 224.5; DB 2; Length 1558;
Best Local Similarity 18.5%, Pred. No. 0.11;
Matches 156; Conservative 123; Mismatches 260; Indels 305; Gaps 29;

QY 44 VKEKKNKDEKRRKDEERKTOEHLKEIMKHIVIEVKGEAVKKAALKLEKVP--- 99
Db 659 IEKLEELHENVLSALENTQSEEEKKEVIDY-----EYKVEVATLIEVEGAE 710
QY 100 ---PSDYLEMYKAIGKIYIVDGDITKHI-----SLEALSE 132
Db 711 ESASITIEIFENLEENAVESINENVAENLEKLNTEVNTVLDKYEEVEISGESLENEM 770
QY 133 DKKKIKDIYGDALLHEHYVAK -EGEVPVLYIOSSE-----DYVENTKAL 178
Db 771 DKAFSELPDNVKGIQENLTLGMPRSITSTVYIOSEKVEVDLENVYSSILDNIEKMEGL 830
QY 179 -----NYYEIGKILSRD-----ILSKINOPYOKFLDLNLTIK 211
Db 831 LNKLENISSTEGVOETVEHVEQNYVDVDPAMKDQGLGLINENGAGLKEKMFNEDFK 890
QY 212 NASDSQODLFTNOLKHPDSEVEFLQNSNEVOEYFAKAFAYIEPQHRDYLQYAP 271
Db 891 SES---DVITVEIKDEP-----VQKEVEKETVSIIEEMENIYDVEE 931
QY 272 EAFNYMDF-----NEQINLSLEELKDQRLSRYEKMEKIK 308
Db 932 EKEDLTDMAIDAVEESLISDSKETESINDKEDVSLVVEVDNDMSSEKVELEK 991
QY 309 QHYQHW-----SDLSSEEGRL-----LKKIQIPLE-----PKK 338
Db 992 NMEELMKDAVEINDITSKLIEPQELNEVADLIKDMEKLEKALSEDSKETIADKD 1051
QY 339 DIHLSQEEKEL-----LKRQIDS-----SDPLSTEE-----KEF----- 370
Db 1052 DTLKRVIEEHDIPTTLDEVVELKQVEDKIEKVSDDLKDEEDILKEVEKIELESEILE 1111
QY 371 -----LKKIQIDIRSLSEEEKELNRIQVDSNPLSEKKEFKLTKIDIQPIQNLQ 426
Db 1112 DYKELKTIEDLIEKKEITEKDHEKF-----EEAEELIKLEADI-----LK 1154
QY 427 DTGGLIDSPSINLVRKQYKRDIONI-----DAL-----LHOSIGSTLYNKIY 469
Db 1155 EVSSLVEEKKLEVEVHLKEVEHILISGDAHINKLEDDLEVPDDLKSGTILDMKGMGE 1214
QY 470 L--YENNMNINLJATLGADVDSTNTKINRGIFNEFKKNFYSSISSNYMIVDINER 527
Db 1215 LGDMDKESLDEVTAKLIGRVSLSLKVLSAAGMEQKTRKKA-----QRPKL 1263
QY 528 DNERLKWRIQSPDRAYLNGKLLQIRNIGLEIKD----- 564
Db 1264 EEVLLKEEVEEPK-----KTKTKKVFYDKDKPEKDEIVEYEMKDEDIDEDIEE 1314
QY 565 -----VOIKOSEKEYIRIDAKVVPKSKIDTK 591
Db 1315 DVEEDIEDKVEDIDEDIDEDIDEDIGDKDEVIDLIYQKEKRIEYAK-----KKLEKK 1371
QY 592 IQEADLINQENKALGLPKYTKLITFNVHNRYASINVESAYILINEMKNINQSDLIKRV 651

DB 1372 VEEGVSGLKKHYDEVW---KYYQKIDKEV-DKEVSKALESKNDYTNVLKGN--QDFFSKY 1425
OY 652 TNTL 655
DB 1426 KNFV 1429

Search completed: December 2, 2001, 13:50:38
Job time: 273 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:46:05 ; Search time 44.8 Seconds

(without alignments)
662.095 Million cell updates/sec

Title: US-09-747-521-2

Perfect score: 4145

Sequence: 1 MNKKKEFIKIVSMKCLVTAT.....KNAPKTPQINDIKFTIINS 809

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4145	100.0	809	1 LEF_BACAN	P15917 bacillus an
2	504	12.2	800	1 CYAA_BACAN	P40136 bacillus an
3	241	5.8	2869	1 RBP1_PLAVB	Q00798 plasmodium
4	235.5	5.7	2663	1 CENE_HUMAN	Q02224 homo sapien
5	234	5.6	997	1 SCPL_RAT	Q03410 rattus norv
6	230	5.5	993	1 SCPL_MOUSE	Q06209 mus musculu
7	223	5.4	794	1 HMR_MOUSE	Q00547 mus musculu
8	218.5	5.3	1075	1 Y124_METJA	Q57588 methanococc
9	218	5.3	756	1 Y128_MYGE	Q49419 mycoplasma
10	216.5	5.2	1679	1 Y109_YEAST	P40457 saccharomyc
11	215	5.2	2245	1 MYSJ_DICDI	P54697 dictyostelli
12	212.5	5.1	1957	1 YDB6_SCHPO	Q10411 schizosacch
13	211.5	5.1	1875	1 MRP1_YEAST	Q00245 saccharomyc
14	211	5.1	1251	1 MRP2_PLAVB	Q00799 plasmodium
15	210	5.1	1939	1 MYH6_MESAU	P13539 mesocricetu
16	208.5	5.0	691	1 Y104_YEAST	P40460 saccharomyc
17	208	5.0	1230	1 SMC3_YEAST	P47037 saccharomyc
18	208	5.0	2710	1 TOXA_CLODI	P16134 clostridium
19	207	4.9	3911	1 AKAG_HUMAN	Q09996 h a kinase
20	204	4.9	1312	1 RA50_YEAST	P12753 saccharomyc
21	204	4.9	1938	1 MYH6_RAT	P02563 rattus norv
22	202	4.9	2748	1 NMU1_YEAST	Q00402 saccharomyc
23	201	4.8	1790	1 USQ1_YEAST	P25886 saccharomyc
24	201	4.8	1935	1 MYH7_PIG	P79293 sus scrofa
25	201	4.8	1940	1 MYH3_RAT	P12847 rattus norv
26	200	4.8	845	1 SCPL_MESAU	Q02566 mus musculu
27	200	4.8	1938	1 MYH6_MOUSE	P08799 dictyostelli
28	200	4.8	2116	1 MYSJ_DICDI	P13533 homo sapien
29	198	4.8	1939	1 MYH6_HUMAN	P30622 homo sapien
30	197.5	4.8	1427	1 REST_HUMAN	P47460 mycoplasma
31	197.5	4.8	1805	1 HMR_MOUSE	P08964 saccharomyc
32	197.5	4.8	1928	1 MYS1_YEAST	P15431 homo sapien
33	196	4.7	976	1 SCPL_HUMAN	

34	196	4.7	1939	1 MYH4_HUMAN	Q9Y623 homo sapien
35	194.5	4.7	1727	1 ALM1_SCHPO	Q9ULK5 schizosacch
36	194.5	4.7	3210	1 CENE_HUMAN	P49454 homo sapien
37	194	4.7	850	1 DP01_ANATH	Q59156 aneroceallu
38	194	4.7	1940	1 MYH3_CHICK	P02565 gallus galli
39	194	4.7	1978	1 MYH8_CHICK	P10587 gallus galli
40	193	4.7	2136	1 YCF2_MARPO	P09975 marchantia
41	193	4.7	2230	1 GOG4_HUMAN	Q13439 homo sapien
42	192.5	4.6	681	1 RPSD_HELPJ	Q9ZMV3 helicobacte
43	191.5	4.6	944	1 NUP1_YEAST	P32380 saccharomyc
44	191.5	4.6	1164	1 BAG_STRAG	P27951 streptococc
45	190.5	4.6	1818	1 HMW2_MYCPN	P75471 mycoplasma

ALIGNMENTS

```

RESULT 1
ID LEF_BACAN STANDARD: PRT; 809 AA.
AC P15917;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE LETHAL FACTOR PRECURSOR (EC 3.4.24.-) (LF).
GN LEF.
OS Bacillus anthracis.
OG Plasmid PX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-49.
RX MEDLINE=90034185; PubMed=2509294;
RA Bragg T.S., Robertson D.L.;
RT "Nucleotide sequence and analysis of the lethal factor gene (lef)
RL from Bacillus anthracis.";
RL (Gene 81:45-54(1989)).
RN [2]
RP SEQUENCE FROM N.A.
RA Lowe J.;
RL Submitted (APR-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RX MEDLINE=95154669; PubMed=7851740;
RT Kochi S.K., Schiavo G., Mock M., Montecucco C.;
RL "Zinc content of the Bacillus anthracis lethal factor.";
RL FEMS Microbiol. Lett. 124:343-348(1994).
CC - FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,
CC AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE
CC DEATH. LEF IS THOUGHT TO BE A LETHAL FACTOR THAT, WHEN ASSOCIATED
CC WITH PA, CAUSES DEATH. LEF IS NOT TOXIC BY ITSELF. PA IS THOUGHT TO
CC BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC CELLS, THEREBY
CC FACILITATING THE INTERNALIZATION OF LEF OR EF.
CC - SUBUNIT: SECRETED ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT
CC PROTEINS, A PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN
CC EDEMA FACTOR (EF). NONE OF THESE IS TOXIC BY ITSELF.
CC - SUBCELLULAR LOCATION: SECRETED.
CC - SIMILARITY: THE PA-BINDING REGION IS FOUND IN BOTH B. ANTHRACIS EF
CC AND LF.
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY M34 (ZINC
CC METALLOPROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M29081; AAA79216.1; -
CC DR EMBL: M30210; AAA22569.1; -

```

PIR: J00032; J00032.
DR MEROPS; M34.001.1.1
DR InterPro: IPR003541; Anthrax_toxina.
DR InterPro: IPR000130; Zn_MTPeptide.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KM HydroLase; Metalloprotease; Zinc; Toxin; Signal; Repeat; Plasmid.
FT SIGNAL 1 33
FT CHAIN 34 809 LETHAL FACTOR.
FT DOMAIN 34 293 PA-BINDING REGION (POTENTIAL).
FT DOMAIN 300 420 REPEATS.
FT METAL 719 719 ZINC (CATALYTIC) (POTENTIAL).
FT ACT_SITE 720 720 POTENTIAL.
FT METAL 723 723 ZINC (CATALYTIC) (POTENTIAL).
SQ SEQUENCE 809 AA; 93786 MW; 8C1684D277310AE CRC64;

Query Match 100.0%; Score 4145; DB 1; Length 809;
Best Local Similarity 100.0%; Pred. No. 1.2e-153;
Matches 809; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKKKFKVISMGLVTAITLSCGPVPIPLVQAGAGHDVGMHYKEREKDKRDEE 60
DB 1 MNKKKFKVISMGLVTAITLSCGPVPIPLVQAGAGHDVGMHYKEREKDKRDEE 60
QY 61 RNKTOEHLKIMKHIVYIEVKGEEAVKKEAEKLEKVPDVLKMYKAIGKITYVDG 120
DB 61 RNKTOEHLKIMKHIVYIEVKGEEAVKKEAEKLEKVPDVLKMYKAIGKITYVDG 120
QY 121 ITHKISLEASDEKRIKIDYGDALLHEHYVYAKEGEPVLIQSSBDYVENTEKALNV 180
DB 121 ITHKISLEASDEKRIKIDYGDALLHEHYVYAKEGEPVLIQSSBDYVENTEKALNV 180
QY 181 YVEIGKILSRILSKINOPYKFLDVLNTIKNASDSGQDLFTNQLEKHPDTSVEFLE 240
DB 181 YVEIGKILSRILSKINOPYKFLDVLNTIKNASDSGQDLFTNQLEKHPDTSVEFLE 240
QY 241 QNSNEVOEVFAKAFAYIEPOHRDVLQYAEAFNYMKFMDQETNLSLEKIDKQRMISR 300
DB 241 QNSNEVOEVFAKAFAYIEPOHRDVLQYAEAFNYMKFMDQETNLSLEKIDKQRMISR 300
QY 301 YEKWEKIKOHYOHMSDSLSEEGRGILKKLQIPPEPKDDIHSLSQEEKELKRIQIDSS 360
DB 301 YEKWEKIKOHYOHMSDSLSEEGRGILKKLQIPPEPKDDIHSLSQEEKELKRIQIDSS 360
QY 361 DFLSTEEREFKLIQIDIRDSLSEBEKELMRIVDSSNPSEKKEFLKKLKDIDQPYD 420
DB 361 DFLSTEEREFKLIQIDIRDSLSEBEKELMRIVDSSNPSEKKEFLKKLKDIDQPYD 420
QY 421 INORLOPTGGLIDSPSTINLDYRKQYKRDIONIDALLHOSIGSTLYNKTYLENNMINLT 480
DB 421 INORLOPTGGLIDSPSTINLDYRKQYKRDIONIDALLHOSIGSTLYNKTYLENNMINLT 480
QY 481 ATTAGADLVSDTNRKINGKIFNEFRKNEKYSISSNYMIVDINERPALDNERLKRIOQSP 540
DB 481 ATTAGADLVSDTNRKINGKIFNEFRKNEKYSISSNYMIVDINERPALDNERLKRIOQSP 540
QY 541 DTRAGYLENGLLIQRNIGLEIKVOIITKQSEKEYIRIDAKVVPKSIDTKIOEQALIN 600
DB 541 DTRAGYLENGLLIQRNIGLEIKVOIITKQSEKEYIRIDAKVVPKSIDTKIOEQALIN 600
QY 601 QEMKALGLPEYTKLTFNPNVNRASNVESAYILNEMKNNISODLKKTYNTLVQDNG 660
DB 601 QEMKALGLPEYTKLTFNPNVNRASNVESAYILNEMKNNISODLKKTYNTLVQDNG 660
QY 661 RFVFTDITLPIAEOYTHODEIYEQVHSGKLYPESRSILLHGSPKGYELANDSEGFHE 720
DB 661 RFVFTDITLPIAEOYTHODEIYEQVHSGKLYPESRSILLHGSPKGYELANDSEGFHE 720
QY 721 FGAAVDDYAGLLDKNOSDLVTSNKKFLIDFKEGSGNLTSGRTNEAEFFAEARLHMST 780
DB 721 FGAAVDDYAGLLDKNOSDLVTSNKKFLIDFKEGSGNLTSGRTNEAEFFAEARLHMST 780
QY 781 DHAERLKVQKNAPKTFQFINDIKFIINS 809
DB 781 DHAERLKVQKNAPKTFQFINDIKFIINS 809

DB 781 DHAERLKVQKNAPKTFQFINDIKFIINS 809
RESULT 2
CYAA_BACAN STANDARD: PRT: 800 AA.
ID CYAA_BACAN
AC P40136;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CALMODULIN-SENSITIVE ADENYLATE CYCLASE PRECURSOR (EC 4.6.1.1) (ATP
DE PYROPHOSPHATE-LYASE) (ADENYLYL CYCLASE) (EDEMA FACTOR) (EF).
GN CYA.
OS Bacillus anthracis.
OG Plasmid PXO1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=69138004; PubMed=2906312;
RA Escuyer V., Duflot E., Sezer O., Danchin A., Mock M.;
RT "Structural homology between virulence-associated bacterial adenylate
cyclases.";
RL Gene 71:293-298(1988).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=89211974; PubMed=3149607;
RA Robertson D.L., Tippetts M.T., Leppla S.H.;
RT "Nucleotide sequence of the Bacillus anthracis edema factor gene
(cya): a calmodulin-dependent adenylate cyclase.";
RL Gene 73:363-371(1988).
RN [3]
RN SEQUENCE FROM N.A.
RA Escuyer V., Duflot E., Mock M., Danchin A.;
RT "Nucleotide sequences expressing adenylate cyclase from B.anthraxis,
RT proteins having the activity of this adenylate cyclase and biological
RT uses.";
RL Patent number EP0366550, 02-MAY-1990.
RN [4]
RN SEQUENCE FROM N.A.
RX MEDLINE=88198021; PubMed=2834337;
RA Tippetts M.T., Robertson D.L.;
RT "Molecular cloning and expression of the Bacillus anthracis edema
RT factor toxin gene: a calmodulin-dependent adenylate cyclase.";
RL J. Bacteriol. 170:2263-2266(1988).
RN [5]
RN SEQUENCE OF 34-48.
RX MEDLINE=89211974; PubMed=3149607;
RA Schmidt J.;
RL unpublished results, cited by:
RL Robertson D.L., Tippetts M.T., Leppla S.H.;
RL Gene 73:363-371(1988).
RN [6]
RN REVIEW.
RX MEDLINE=93119764; PubMed=8418825;
RA Danchin A.;
RT "Physiology of adenylate cyclases.";
RT Adv. Second Messenger Phosphoprotein Res. 27:109-162(1993).
CC -1- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,
CC AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE
CC DEATH. EF IS A CALMODULIN-DEPENDENT ADENYLYL CYCLASE THAT, WHEN
CC ASSOCIATED WITH PA, CAUSES EDEMA. EF IS NOT TOXIC BY ITSELF. PA IS
CC THOUGHT TO BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC CELLS,
CC THEREBY FACILITATING THE INTERNALIZATION OF EF OR LF.
CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC -1- ENZYME REGULATION: ACTIVATED BY HOST CALMODULIN.
CC -1- SUBUNIT: ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT PROTEINS, A
CC PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN EDEMA FACTOR
CC (EF). NONE OF THESE IS TOXIC BY ITSELF.
CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-2 FAMILY.
CC -1- SIMILARITY: THE PA-BINDING REGION IS FOUND IN BOTH B.ANTHRACIS EF

CC AND LF.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M23179; AAA22374.1; -
 DR EMBL: M24074; AAA9215.1; -
 DR EMBL: A07289; CAA00652.1; ALT_SEQ.
 DR InterPro: IPR003541; Anthrax_toxinA.
 KW Lyase: CAMP synthetase; Toxin; ATP-binding; Signal; Plasmid.
 FT SIGNAL 1 33
 FT CHAIN 1 33
 FT DOMAIN 34 800 CALMODULIN-SENSITIVE ADENYLATE CYCLASE.
 FT DOMAIN 34 288 PA-BINDING REGION (POTENTIAL).
 FT NP_BIND 289 680 CATALYTIC.
 FT NP_BIND 347 354 ATP (POTENTIAL).
 FT CONFLICT 350 350 V -> E (IN REF. 2).
 FT CONFLICT 510 510 Q -> T (IN REF. 2).
 FT CONFLICT 512 513 EM -> RM (IN REF. 2).
 FT CONFLICT 760 760 V -> L (IN REF. 3).
 FT CONFLICT 760 760
 FT SEQUENCE 800 AA; 92477 MW; F4F7EB485DF4C5A6 CRC64;

Query Match 12.2%; Score 504; DB 1; Length 800;
 Best Local Similarity 24.1%; Pred. No. 4.9e-13;
 Matches 214; Conservative 154; Mismatches 297; Indels 224; Gaps 40;

QY 4 KKEFI-----KVISMGLVATITLSPVFIPLVQAGAGHGDCVMHYKERRKNKDENRKME 59
 DB 3 RKKEFPKFSITISFVLLPAISSQAIEYNA-----MNEYTESDIKRNHKT 49
 QY 60 ERNKTOEHLKEMKHYKIEYKGEAVYKKEAEKLEKPEPSDYLEMYKAIGKIYVDG 119
 DB 50 EKNTKEKEKFSINNVLVTEFTETLTKIQOTDLKKIPROVLEIYSELGGEIYFTDI 109
 QY 120 DITKITSLEALSEDKKIKIDYKQALHLEHYAKEGEPVLYVIOSSDYVENEKALN 179
 DB 110 DLVEHKELODSEEEKNSNSRGEVFPASRFVEKRETPKLIINIDYVAINSSOSKE 168
 QY 180 VYEELGKILSPILSKINQPYOKPLD-VLNTIKNAS-DSDQDILFTNOLAE---HPT 232
 DB 169 VYEEIGKIGSLDIISK-----DKSLDPEFLNLIKSLDSDSDSLFSCKFEKLELNK 223
 QY 233 DESVEFLDONSNEQVEVPAKAFAYIEPOHRDVLQIYAPAEFNYMDKFNEDINLSLEL 292
 DB 224 SIDNFIKENLTFQHAFLAFSYFAPDRIVLELYAFDMFEYMNKL----- 271
 QY 293 KDQRLSLREK-WEKIKOHYHMSDLSSEGRGLIKLQIPIEPKDDIHSLSOEKE 350
 DB 272 -----EKGGFEKI-----SESLKEG-----VEKDRIDVL---KGEKA 301
 QY 351 LKRIQIDSDPLSTEKEKFLKQIDIRDSSEBEKELNLIQVDSNPLSEKEFEFLK 410
 DB 302 L-----KASGLVPEHDAFKKI-----ARELNTYILFEPVKLATN---LI 339
 QY 411 KLIKIDIQPIYINQRIQDGLIDSPSINDVVKQY-----KRDIONIDALL-HQ-SIG 461
 DB 340 KSGVATKGLNIVKSSDMQPVAGIYIPFODDLSKKHQOLAVKGNLENKKSITHEHGELG 399
 QY 462 STLYNKKIYL-YENMININLTATLGADLVDSNTKINRGIFNEFKNFK-----YSISSN 515
 DB 400 -----KIPKLDLRIIEELK-----ENGIIILKKEKKEINDGKKYLLLEEN 438
 QY 516 YVIIVINERPALDNERLKRIQLSPDTRAGYLE-NCKLLQNIIGLEIYDVO-IIGQSEK 573
 DB 439 NOVYE--RISDENNEVOYK-----TKGKITVLEGEKFMWRNIEYMANVEGVLPPLTA 490
 QY 574 EXIRIDAKVVPKSIDTKIQEQAQLNINQEMNKALGLPKYTK-----LITFVNINRFA 625

DB 491 DY---DLFALAPSLTEIRKOIPQ-----KEMDKVYVNTPNLSLEKQGVYLLIKYGERK-- 541
 QY 626 SNIVESAYLILNKMKNINIOSDLIKKTYNTLVDSNGRFVFDITLFLINAEYTHOD--EY 683
 DB 542 ---PDSTKCTLSWQKOMDRL-----NEAVKTYGYGVGVNHTGSDNEDEFP 587
 QY 684 EGVHSGLVYPEERSILLHG---PSKGVLELNDSEGFIEFGVADYDA-----GYL----- 732
 DB 588 EKNELEFIINPEEFLITKNWEMTGRIENKITGKQIYLYFNNSYKKAIAQKNAIYEMD 647
 QY 733 -LDKNOSDLVTNSKFTI-----DIFEEGSSN-----LTSYGRT 764
 DB 648 PIKKAKINPTSAEFKIKNLSIRSSNVCYKSDGKDEFAKESVKKIAGLYSDYNS 707
 QY 765 NEKEFFAE-----AFRLMSTDAER-LKVQKNARK---TFPINDOI 803
 DB 708 ANHIFOEKRRKISIRGIAVNEIENVLKSKQIADEYKNKYFOLKERI 756

RESULT 3
 ID RBPL_PLAVB STANDARD; PRT; 2869 AA.
 AC Q00798;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
 GN RBPL.
 OS Plasmodium vivax (strain Belem).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=31273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92315338; PubMed=1617731;
 RA Galinski M.R., Medina C.C., Ingavalle P., Barnwell J.W.;
 RT "A reticulocyte-binding protein complex of Plasmodium vivax
 RT mercoites";
 RL Cell 69:1213-1226(1992).
 CC -I- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
 CC HUMAN RETICULOCYTE CELLS.
 CC -I- SUBUNIT: HOMODIMER (POTENTIAL).
 CC -I- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M88097; AAA29743.1; -
 DR HSSP: P36956; IAM9.
 KW Malaria; Receptor; Signal; Transmembrane.
 FT SIGNAL 1 17
 FT CHAIN 1 17
 FT DOMAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
 FT DOMAIN 18 2807 EXTRACELLULAR.
 FT TRANSMEM 2808 2826 POTENTIAL.
 FT DOMAIN 2827 2869 CYTOPLASMIC.
 FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
 FT SEQUENCE 2869 AA; 330213 MW; B9DBE442205BECFF CRC64;

Query Match 5.8%; Score 241; DB 1; Length 2869;
 Best Local Similarity 18.9%; Pred. No. 0.032;
 Matches 194; Conservative 161; Mismatches 326; Indels 346; Gaps 46;

QY 39 DVGMYKKEKKNKDENR-----KDEERNKTOEHLKETIMKIHYKI----- 79
 DB 746 DIALIEVEKEFTENKKESTLEMLKDEE---MEKKLODAKETFALKNFVSDDKLTDVYT 801
 QY 80 -----EVKGEAVYKKEAEKLEKPEPSDYLEMYKAIGKI-----YIVDD-ITKHI 125


```

OY 82 KGEAVKKEAEKLEKVPDVLNEMKAIGKIYIVDQITKHISLEALSEDKKIK--- 138
DB 584 LRE-----KKEQIKKLEQYIDSKLENIMK-----DLS--YISLEST-EDPKQMKOVL 627
OY 139 -----DIYKDALLHEHYVAAEGEYEPVLVIOSSSEDEVVENTEKALNYYEIGKILSR 190
DB 628 FDAETVALDARKESAFILSENLELEKEMKE---LATYIKOMENDIQLOQSELEAKKKMOV 684
OY 191 DILSLINOPYOKFLDVLNLIKASDSG---ODLFTNOLKEHPDPSVEF---LEQNSN 244
DB 665 DLEKELQSAFNITRLTSLI-----DGKVPDDLCLNLEKGIITDLOKELKEVEENEA 738
OY 245 EKOGEF-----AKAFAYIIEPOHRDY-----LOLAPAEAFNMDKF-----NEQDINLS 288
DB 739 LREYIILSELKSLPSEVRLRKEIQDKSEELHIITSE---KDLFSEVYKESVQGL 794
OY 289 LEEL---KDQRLMSRYEKWEKIKOHYQWMSDSLSEGRGLKLOIPIPKKDDIHS-- 343
DB 795 LEIIGTKTD-LATQSNYKSTDOEFQNF-----KTLHMDFOQYKMWLEENE 841
OY 344 -LSQEKELKRIQ-IDSS-DPLSTE-----EKFLKQLQIDTD 380
DB 842 RMOQEVINLSKEAQKFDSSLGALKTELSTYQLOEKTREVOERLNMEDQKE-QLEND 900
OY 381 S---LSEEEKELNRIQVDSNPLSEKEKFLKLLIDQPYDINQRLQDT----- 428
DB 901 SPLQYVERKTL-----ITEKLOQTLSEYKVTLQOEKDDLKQLOESIQIENDQK 949
OY 429 GGLIDSPSINLDVROKYKRDIONDALLHQSIGSTLYNKIYIENMNINLTATIGADLV 488
DB 950 SGIHDTVMNMIIDTQELRMALESLEKQ--HQETINFLKSISESVSNHL-----ME 998
OY 488 DSTDMTKIRNGIFNEPKKFKYSISSNMYIVDINERPALDNERLKRIOLOSPTRAGIYE 548
DB 999 EMTGETK-----DEFQOK-----MVGIDKKQDLEAAN----- 1025
OY 549 NGKLLIQRNIGLEIKDVQITKQSEKEYIRIDAKVYPSKSIDTKIOE-AQLINQEWNKAL 607
DB 1026 -----TOTLTDADVNDNEITEBOQRKIFSLQEKNELOQMLESVIAKEBQKLDKKN--- 1076
OY 608 GLPKYTKLTFFNVHNRKYNIVESAVYLINENKKNIOQLIKVTVLYLVGNGRFVETDI 667
DB 1077 -----IEMTINQ-----EELRLLDGLDELK--QOEIVAOEKHNAIKKEGELSRTCD 1120
OY 668 TLPNIAEOYTHODEIYEOVHSGKGLVYPSRSILLHGPSKVELRNDSEGFIEHFGAVVD 727
DB 1121 RLAEVEEKLEKESQOLQEQOQLLVQEMSEMOKKINEIENLKNELKKELTLEHMETE 1180
OY 728 --YAGILDKNOSDL--VTNSKRFI---DIFKEEGSNLTSYGRTEAEFEFAFRL--- 776
DB 1181 RELQAKLKNENEEVAKSTIKERKVLKELOKSFETERDHLRGYIREIAVGLQTKKEELKIA 1240
OY 777 -MHSIDHAERLKVQKN--APKTFQFIINDQ 802
DB 1241 HILKEHOETIDELRVSSEKTAQIINTQ 1269

RESULT 5
SCPI_RAT STANDARD: PRT: 997 AA.
AC 003410:
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SYNAPTONEMAL COMPLEX PROTEIN 1 (SCP-1 PROTEIN).
GN SYCP1 OR SCPI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 11
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=Testis;
RX MEDLINE=93099884; PubMed=1464329;
RA Meuwissen R.L.J., Offenberg H.H., Dietrich A.J., Riesewijk A.,
RA Terseel M., Heyting C.,
RT "A coiled-coil related protein specific for synapsed regions of
RT meiotic prophase chromosomes.",
RL EMBO J. 11:5091-5100(1992).
CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF
CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
CC CHROMOSOMES DURING MEIOTIC PROPHASE.
CC SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
CC SYNAPTONEMAL COMPLEX.
CC -1- TISSUE SPECIFICITY: TESTIS.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EXCLUSIVELY IN MEIOTIC PROPHASE
CC CELLS.
CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
CC DOMAIN HAS DNA-BINDING CAPACITY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
CC CORRECTED IN POSITION 6 TO MAXIMIZE THE SIMILARITY WITH THE
CC OTHER SPECIES SYCP1 SEQUENCES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC EMBL: X67805; CAA8006.1; ALT_FRAME.
CC DR PIR: S28061; S28061.
CC KW Nuclear protein; Meiosis; Cell division; Phosphorylation;
CC DNA-binding; Coiled coil.
CC FT DOMAIN 108 819 COILED COIL (POTENTIAL).
CC FT DOMAIN 118 121 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT DOMAIN 701 704 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT DOMAIN 902 905 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT DOMAIN 982 990 ARG/LYS-RICH (BASIC).
CC FT SEQUENCE 997 AA; 116511 MW; 229D59823FD684BE CRC64;

Query Match 5.6%; Score 234; DB 1; Length 997;
Best local similarity 21.7%; Pred. No. 0.017;
Matches 199; Conservative 158; Mismatches 349; Indels 210; Gaps 50;

OY 44 VKKEKKNQKDKKDEERNKTOEHLKEIMKHIVKIE-----YKGEAVK----- 88
DB 128 LKQKKNKLOENKRIITEQRAIOELOFENKESVLAEEIOENKKDLKENNATRHMCNLL 187
OY 89 KEAAERLLEKVPDVLNEMKAIGKIYI--VDGDIYTHI-----SLFALSEDKK-----KIK 138
DB 188 KETCARSAEK--TSKYEYEEETROYVVDLNNIEMKILAFEEELRYQAEARLEMEFKIK 245
OY 139 DIYKDALLHEHY-----YVAKGEYEPVLVIOSSSEDEVVENTEKALNYYEIGKILSRDILSK 195
DB 246 EDHEKIQHLEEEYQKEVNNENOVSLILIOSTER--ENKKMDLTFLE-----ESRD---K 296
OY 196 INOPYOKF-----LDVLNLIKASDSGODLFTNQ-----LKEH---PTDSVEF 238
DB 297 ANOLEKTKIQDENLKELNKKDHLTSELBDIMQMSKSTOKTLEEDLIQITKTIYQL 356
OY 239 LEQNSNEVOEV-----FAKAFAYIIE-----POHRDVLQLYAPE----- 272
DB 357 TEEKKQOMELNKAATKTHSLVYELKATCTLEELRTEDQGRLENNEDDQKLTLMELQK 416
OY 273 --AFNYMDKF--NQEINLSLEELK-----DQRLMSRYEKWEKIKOHYQWMSDSL---EE 321
DB 417 SSELKMTKFKNNKEV--ELEELKTLIAEDQKLLDEKQVEKLAELQEGQELTFLQOT 474

```

```

OY 322 GRGLKKQIDPIEPRKODIHSLSOE-----EKELIKRIQID-SSDFLSTEEKF--- 370
D 475 REKEIHLEVOVYTKTSEEHYLKQVEEMKTELEKKNIELNANSMDLLENKKLVQOE 534
OY 371 -----LKKQIDIRSLSEEEK-----ELNRIQVDSNPISKEKEFLK-----KTK 413
D 535 ASDVAVLELKKHEDIIINCKQEEPMKQIETLEEKEMLRRELSVREFTQGDDEVYCK 594
OY 414 LDIO-----PYDINORLQDTGLIDPSINLDPVKQYKRDIONIDALHQ-----SIG 461
D 595 LDKSEENARSIIEYVLK--EKQMIKENKNL--KKQIEKKSNIIE-LHQENKALKKK 650
OY 462 STLYNK-ITYLENNMINNLTATLGADLVSDTNKIN-RGIFNEPKKNFKY-SISSNMI 518
D 651 SSANCKOLMAE-ITVKNLELELA-----STKQFEEMINNYOKEIEIKISEEKL 701
OY 519 VDIENERPALDERLKWRIQTSPTFRAGY-----LENGLLIQIRNGLEIKVOIITKQ 570
D 702 GEVKKAKATYDEAVK--LQKEITLRCQKIAEMVALMEKHQYDKIVEERDSEGLYKN 759
OY 571 SEKEVIRIDAKVYPSKIDTKIQE-----AQLNINOENKALG-PRYTKLITFNVHRY 624
D 760 REGE--QSAKVALETETLSNIRNELVSLKKQLEVEKEKEKLMQENTALTITKKKKKI 817
OY 625 ASNIVEASVYLINEMK--NNIQSDLIKVTNYIVDNG-----RPFVTDITLPNIAE 674
D 818 QASLLESPE--ATSMKFDSTPTSPONISIRLSSMSDGSKSDNDRSLASAKISLTIVTK 875
OY 675 QYTHQDIEVEVHSGKGLYVPSRSILHGPCK-----GVELRNDSE-----GFINEFG 722
D 876 EYT-----VTPPKTSYQREPNKTLPPGSSKKKKRYTFEPFVNSDSSSETTDLISLVE-- 928
OY 723 HAVDYAGVLLDKNQSD--LVTSNKKETIDIFKEEGSNLTSYGTNEAEFEAFERLMHST 780
D 929 ---EDISNRIYNNNTPSHLLVTKPTKQPLSLSTPASTTKGSLKMR--EDRWATIAKI 983
OY 781 DHAERLKVOKNAPKTF 796
D 984 DRRRL--KEAEKLIF 996

```

```

RN [4]
RP SEQUENCE OF 95-787 FROM N.A.
RC STRAIN=ICR; TISSUE=Testis;
RA Tsuchida J., Nishina Y., Nozaki M., Uchida K., Nishimune Y.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF
CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
CC CHROMOSOMES DURING MEIOTIC PROPHASE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
CC SYNAPTONEMAL COMPLEX (BY SIMILARITY).
CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z38118; CAAB6262.1; -
DR EMBL: LA1069; AAA64514.1; ALT_INT.
DR EMBL: U62864; AAC53335.1; -
DR EMBL: U62860; AAC53335.1; JOINED.
DR EMBL: U62861; AAC53335.1; JOINED.
DR EMBL: U62862; AAC53335.1; JOINED.
DR EMBL: U62863; AAC53335.1; JOINED.
DR EMBL: D88539; BAA13639.1; -
DR TMD: MGI:105931; Sycp1.
KW Nuclear protein; Meiosis; Cell division; Phosphorylation;
KW DNA-binding; Coiled coil.
FT DOMAIN 12 97
FT DOMAIN 104 815 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 114 117 COILED COIL (POTENTIAL).
FT DOMAIN 697 700 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 898 901 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 978 986 ARG/LYS-RICH (BASIC).
FT CONFLICT 527 527 F -> L (IN REF. 2).
SQ SEQUENCE 993 AA; 115962 MW; 1AAFA790D64FAFE6 CRC64;

```

Query Match 5.5%; Score 230; DB 1; Length 993;
 Best Local Similarity 21.8%; Pred. No. 0.024;
 Matches 202; Conservative 149; Mismatches 346; Indels 228; Gaps 50;

```

OY 44 VKEKKKNKDEKRRKDEFRNKTOEHLKEIMKHVYKIEVKGGEAVK----- 88
D 124 LKQENKLNQEKRIITEQORKA-----IOELQENKVSILKEEIQENKDLKENNATIIHW 179
OY 89 -----KEAAEKLLKVPDVLVEMVKAIGKITYI-VGDITRKHI-----SLEALSDEKK--- 135
D 180 CNLKEKFCARSAEK--TNKYEYEREETROYVDLNSNIEKKIILFAEELRVAQAEVARLEMH 237
OY 136 -KINDIYKDALLHEHY---VYAKEGEPVLYIOSSSEDIYVENTKALNVYIEIKIISRD 191
D 238 FKLEDEHKIOHLEEEYQKEVNNKENOVSELLIQAER--ENKKKDLTFPLE-----BSRD 291
OY 192 ILSKINQYQKF-----LDVLNTRKNASDSDGDL-----LFTNQLKEHPTDFVEF 238
D 292 ---KANOLEETKTIQDENIKELSKKDHLSLEDIDKMSQRSSTOKALEEDQITKT 348
OY 239 LEQNSNEVQE-----VFAKAFAYIE-----POHRDVLQDYAP 271
D 349 ISQ-LTEVKEAQEMELKNKAKTTHSFVYTELKATCTLEELLRTFQQRLEKKNEDQLITV 407
OY 272 E-----AFNVMKDF-NEQELNLSLEELK-----DQMSLSYEKWEKIKQHYQMSWSLS 319

```

```

Db 408 ELQKSNLEEMTKFNKNKV--ELEELKNILAEQDKLDEKQVKEAELEQEKDEELT 465
Oy 320 ---EEGRGLKQIPIEPKDDIHSLSOE-----EKELKRIQIDS--DELSTEEK 368
Db 466 FILLEPEKEVHDLQEVYATKSEQHYLKQVEEMKTELEKEKIKNTLTAACMLLENK 555
Oy 369 EF-----LKLQIDIRDSLSEEEKELNRIQVDSNPLSEKE-----KE 407
Db 526 KFOVASDMALFKKHQEDINCKQOER-LKQIE-----NLEEKEMHLRDELSVRKE 579
Oy 408 FLK-----KLKIDIPYDINQ-----RLQDTGGLDSINIDVKKQKRIQNTDALL 456
Db 580 FIOQGEVCKKLDKSEENRSIECEVLKKEKQKILKSCNNL--KQOVENKSKNTEE-L 636
Oy 457 HOSIGSTLKNK-----LYLVENMNINULTAGADLVSTQNTKINGIFENEFKNKF 509
Db 637 HOB-KMTLKKSASAEIKQLMAYE-IVSKLELEL-----ESTQKOR--FEEMTNNTQ 663
Oy 510 ----YSISSNMYVIDINERPALDNRLKWRIOQSPDTRAGY-----LENGKLILQK 556
Db 684 KEIENKRISEGLKGEVEKAKATVDEAVK--LQKEIDLQOQHKIABVALMEKHKKHQYK 741
Oy 557 NIGLEKNDVOIIKQSEKEITRIDAKVPRKSIDTKIQE-----AQINQOENKALGLPK 611
Db 742 IVEERDSEGLKYKNRQOE--QSSAKTALLETLSNINRELVSLKQKLEIEKEEKLKMAK 799
Oy 612 -YTKLITFVHNRYASNIYESAVLLENK---NNIQSLIKKVTYVLVDG-----NG 660
Db 800 EMTAILKDKKDKKIQASLESPE--ATSWKFPDSTKPPSONISLSSSMQSGSKKDRNDL 857
Oy 661 REVFTDITLPNIAEOYTHODEIYEQVHRSK-GLYVPEPSRSLHGPBGVLELRDSE---- 715
Db 858 RASAKSILFTYTKETVTPPKKSITYQRENKXIPFGSGNKKKKKTFEDVNDSSSETAD 917
Oy 716 --GFHEFGHAYDVAGYLLDKNQSD--LVYNSKFRIDIFKEGSGNLTSGRTNEAEFFA 771
Db 918 LLSLVSE-----EDVSNRLYDNNPPDSHLVTKPTPLSLSTPASPMKSGSLKKMR--E 970
Oy 772 EAFRLMHSSTDHMRLEKVKONAPKTF 796
Db 971 DKMTTIAKIDRRRL--KEAEKLF 992

```

```

RX MEDLINE-99107769; PubMed-9889313;
RA Fieber C., Plug R., Sleeman J., Dall P., Ponta H., Hofmann M.;
RT "Characterization of the murine gene encoding the intracellular
RT hyaluronan receptor IHABP."
RL Gene 226:41-50(1999).
[4]
RN
RP SEQUENCE OF 164-794 FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN-BALB/C; TISSUE-Fibroblast;
RX MEDLINE-96011639; PubMed-7590272;
RA Entwistle J., Zhang S., Yang B., Wong C., Li Q., Hall C.L., A.J.,
RA Kowat M., Greenberg A.H., Turley E.A.;
RT "Characterization of the murine gene encoding the hyaluronan receptor
RT RHAMM."
RL Gene 163:233-238(1995).
[5]
RN
RP SEQUENCE OF 318-794 FROM N.A., AND CHARACTERIZATION.
RC STRAIN-BALB/C;
RX MEDLINE-92299690; PubMed-1376732;
RA Hardwick C., Hoare K., Owens R., Hohn H.P., Hook M., Moore D.,
RA Crisps V., Austen L., Nance D.M., Turley E.A.;
RT "Molecular cloning of a novel hyaluronan receptor that mediates tumor
RT cell motility."
RL J. Cell Biol. 117:1343-1350(1992).
[6]
RN
RP CHARACTERIZATION.
RX MEDLINE-94308286; PubMed-7518470;
RA Hall C.L., Wang C., Lange L.A., Turley E.A.;
RT "Hyaluronan and the hyaluronan receptor RHAMM promote focal adhesion
RT turnover and transient tyrosine kinase activity."
RL J. Cell Biol. 126:575-588(1994).
[7]
RN
RP ERK REGULATION, AND SUBCELLULAR LOCATION.
RX MEDLINE-98225222; PubMed-9556628;
RA Zhang S., Chang M.C., Zylka D., Turley S., Harrison R., Turley E.A.;
RT "The hyaluronan receptor RHAMM regulates extracellular-regulated
RT kinase."
RL J. Biol. Chem. 273:11342-11348(1998).
[8]
RN
RP REVIEW.
RX MEDLINE-99059494; PubMed-9845361;
RA Hofmann M., Assmann V., Fieber C., Sleeman J.P., Moll J., Ponta H.,
RA Hart I.R., Herrlich P.;
RT "Problems with RHAMM: a new link between surface adhesion and
RT oncogenesis?"
RL Cell 95:591-592(1998).
[9]
RN
RP -1- FUNCTION: INVOLVED IN CELL MOTILITY. WHEN HYALURONAN BINDS TO
CC HMHR, THE PHOSPHORYLATION OF A NUMBER OF PROTEINS, INCLUDING THE
CC FOCAL ADHESION KINASE OCCURS. MAY ALSO BE INVOLVED IN CELLULAR
CC TRANSFORMATION AND METASTASIS FORMATION, AND IN REGULATING
CC EXTRACELLULAR-REGULATED KINASE (ERK) ACTIVITY.
CC -1- SUBUNIT: SUBUNIT OF THE HARC COMPLEX.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE AND CYTOPLASMIC.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; RHAMM1V4 (SHOWN HERE) AND
CC RHAMM1; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBIOUITOUSLY EXPRESSED.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF031932; AAC12655.1; -
DR EMBL; AF079222; AAD08670.1; -
DR EMBL; AJ005919; CAA06768.1; -
DR EMBL; AJ005920; CAA06768.1; JOINED.
DR EMBL; AJ005921; CAA06768.1; JOINED.
DR EMBL; AJ005922; CAA06768.1; JOINED.
DR EMBL; AJ005923; CAA06768.1; JOINED.
DR EMBL; AJ005924; CAA06768.1; JOINED.
DR EMBL; X64550; CAA45849.1; -

```



```

OY 66 EEHLKEIMKHIVKEVGEAAVKAFAE-----KLEKVPDVLDM 106
DB 323 DSDL-DLSKPIEELIKEDLKKEGKIKGLDILLIILAFYLKHEKNP-----DE 377
OY 107 YKAIGKITYVODITKHLSLEASDDKKIKDIYKOLLEHHYAAKEGEPVLVLOS 166
DB 378 YK-----KHI-----ENLKLKD-----K 393
OY 167 SDDVTEKALNYYEIGKILSDILSKINPYOKFLDVLNITKNASDDGDLFTNG 226
DB 394 KEETLIN-----LONIKRHLILLIDEHNRQYIGLGMKRTTPNATFEGTGT 443
OY 227 --LKEHPDVSVEFLQNSNEVEVFAKAFAYIEPHRDVLOLYPEAFNVMK--FNE 282
DB 444 PVFKNEKNFT--EFSYPEKGEFY-----LDVYFIGDSINDKFTL--PLYQYVKEEDINS 495
OY 283 QINISLEFLKQORMLSYKEWEKKIYOHV--SDLSSEGRGL-----KKIQPIE 334
DB 496 EGIQITLDE-----EDIEKFIDEWIKRGEDINLFDKRLPKYINKSKITILN 542
OY 335 PKDDIHSLSOEKELLKRIQIDSSD-----LSTEE--KEFIKQIDIRDSLEE 385
DB 543 PRIDRY-----AKIYDRIEDIEDENEFKAMAVAVNRLGCVREFKALDKYIKKEGDE 596
OY 386 EKELLNRIOVSSNPLSEK--EFLKRLKLDIOPYD--INORLOD-----427
DB 597 AKKMAEVVYTHHN--EKEKELIEYMKLKEKERNSDNEINQIREFLSSENKILL 653
OY 428 -----TGGLDPSPIN--LDVKQYKRDIONID-----ALHOSIG--STLY 465
DB 654 VTMMLTGT--FDAPRLKAVYLDKPLYGHRLLQALATNRPYDKERGLIVDSGLKFLVLT 711
OY 466 NKIYLYENNINNLATLGADLVSDTNKIRGTFENK-----KNKYSSISN 515
DB 712 EFMALYNMLAEELIRDEFKNNLISIDE-----IFOEKLELVKESLKNLK--INDE 763
OY 516 YMIIVDINERPAL-----DNERLKWRIQLSPDTRAGYLENGK-----LILQ 555
DB 764 DLSIDVNTLKTLLKKNDFNNNELKEL-----DLIAFYADGKNARILKLDLKAIVIKLY 819
OY 556 RNIG-----LEIKDVOIKQSEKEYIRIDAVVPKSIDTQIOEA-----OL 597
DB 820 KALGSYPOKIFYIEDIELLSFT--YAYLIKRLKPKKSNRKFWEELISFIHNKMLVDL 876
OY 598 NINQEMNNAIGLRYTKLTTFVHNRYASNIYESAILINEMKNNI--QSDLIKAVTVY 654
DB 877 TVIEEIN--LNPDDLDKLIKENIGKREIKRAVANYFYL--KNSILDRQHDPIYK---- 927
OY 655 IVDGNGREFVDTILPNIAE--QYTHODEIYGVHSGLYPESRILHGPSKGVLELN 712
DB 928 -----EILERLERLRQWIMKRIKD-----IYNAIKNLMELKN 962
OY 713 D-----SEGFIHERGHAVDAGYLLDKNSDL--VTNSKKFIDIFEKESNLTSYGR 763
DB 963 NYDKRIKGSISIERIKESISTYIGENILKQD--DIKINLENTKELI--TKMQNLKLSK 1017
OY 764 TNEAEFFAE 772
DB 1018 LQKKFKKE 1026

```

```

RESULT 9
V328_MYCGE STANDARD: PRT: 756 AA.
AC Q49419; Q49308; Q49320;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MG328.
GN MG328.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
Mycoplasmataceae; Mycoplasma.

```

```

OX NCBI_TaxID=2097;
RN [1]
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bolt K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of mycoplasma genitalium.";
RT science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 19-113 AND 155-272 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bolt K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RT J. Bacteriol. 175:7918-7930(1993).
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
CC EMBL: U39714; AAC71552.1; -
CC EMBL: U02203; AAD12492.1; -
CC EMBL: U02188; AAD12472.1; -
CC TIGR: MG328; -
CC KW Hypothetical protein; Complete proteome.
CC FT CONFLICT 52 52 T -> R (IN REF. 2).
CC FT CONFLICT 65 65 S -> W (IN REF. 2).
CC SQ SEQUENCE 756 AA; 88407 MW; 1D3AD4EFOAB5C068 CRC64;

```

```

Query Match 5.3%; Score 218; DB 1; Length 756;
Best Local Similarity 21.6%; Pred. No. 0.051;
Matches 181; Conservative 145; Mismatches 268; Indels 244; Gaps 49;

```

```

OY 46 EKEKN-----KDEKRRKDE-----ERNKTOENHLEIKKHIVKEVGEAV 87
DB 61 EKEKSTVSEFNNMIDLKVLKDKWKNVNLHNYANQLKTRREEOQTUQTMAFOEVDQSYL 120
OY 88 KKEAAKLEKVP-----DYLEMKKATGKLYIVDGTKHISLEAL-----SPD 133
DB 121 TKEIKTFQELKPSVTTAEDDKDE-----IKPEKTKOVSEELFNOPSEINET 169
OY 134 KKKIKDIYKDALHBYVYAKGEYEPVLVLOSSEY--VENEKALNYYEIGKILSRD 191
DB 170 KKEVQVIFSPDKV-----KEP-----EQFDFYSIENTKALN--PVNKTITQVD 211
OY 192 ILSKINQPY--QKFLDVLNITKNASD--GODLLFTN--QLKEHPTDSVEFLQNSNE 245
DB 212 --QNDQDPVYVRIKLEQHPYTKKVDLDYNNKELLLENADLKK-----QIDDLKENMND 264
OY 246 VQEVFAKAFAYIIEPHRDVLOLYAPEAFNVMKFNQEOENLSEELKQORMLSYKEWE 305
DB 265 --QIF-----DL-----EQET-----DCLK--RLLSB---E 283
OY 306 KIKQ--HYQMSDSLSEEGRLKLLKLO--IPIEPKRDIDILHSLS--QEEKELKRIQIDS 359
DB 284 KSKHLHTKKLQDOLLQENNDIYEQLOKNKRYAIPLSDEVNEELNKQKALLSD--QLDA 342
OY 360 SDFLSTEEKELKLDIIRDSLSEEEKELNRIQVSSNPLSEKKEFLK--KKLDI 416
DB 343 LKNKSSNVQOOLALPLV--LNNQINELQONOLLTAREANORLIDVEQENDFLKNELKHLN 401
OY 417 QPDIYQRIQDIDG-----LIDSPSINLDVRYKQYKRIQIONIDALHOSIGSTLYLNKLYLE 472

```

```
Db 402 TSNDENEYDLDLNOYELLPBENETKPD-KIOVQOALNLD---YOKTSLKHE----- 452
Oy 473 NMNINNTATIGADLVSTDFKINRGIFNEFKKFKSYSSNMYMVIDINERPALD----- 528
Db 453 -----NDV-----LIDIEWTRSKDNDFNMTKNSFE-----EQKKALDEKLN 489
Oy 529 -----NEMLKRI-QASP-DTRAGYLENGKLLIQNRIGLEIKDVOILIKOSEKEIRIDAK 581
Db 490 GLTIQNOOLQOKIAIELEBWEENEKSNLNTNOLV---NLQOOLKDSQNLFN----- 535
Oy 582 VVPSKIDTKIOEAOALNIOEMN---KALGPKYTKLITFNVHRNRYASNIVESAYLTIN 637
Db 536 -VAODKLAT-LEEVNLAINEKINLEDELSG-----SENSNNLAK--LOADHETLO 583
Oy 638 EMKNNISDLIKKYTNVLYVNGRFEVFTDITLPNIAEQYTHODEIYEQVHSGKGLVPESR 697
Db 584 ESYGKLKTDPEKTKKKNKLNDA-----EYOQDLSAFEEETNSE---LEKAK 626
Oy 698 SILHGSPKGYELR---NDSEGFIERGHAVDDYAGYLLDKNOSDLYVNSKKFDIFKEE 754
Db 627 QSLASDSENNOLKQOINSLNARKEL---QETPVTSDHDELLET---LKIEKEQ 676
Oy 755 GSNLTSGRTNEAEFFAFRLMSTHDAERLKVKOKNAPKTFEOFINDQIK---FLINS 809
Db 677 -LFIENQALQOQLOVFNDI-----SANOTEIKASDEKDEVEIKPKIKKRDVEYON 728

RESULT 10
ID Y109 YEAST STANDARD; PRT; 1679 AA.
AC P40457;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE HYPOTHEICAL 195.1 KDA PROTEIN IN DNA43-UB11 INTERGENIC REGION.
GN Y1149C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Goules S., Hamlyn N., Horsnell T.S., Hunt S., Jagsels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z38059; CAAB6129.1; -.
DR PIR: S48385; S48385.
DR SGD: S000141; MUP2.
KW Hypothetical protein.
SQ
SEQUENCE 1679 AA; 195141 MW; 298950CC52202D8F CRC64;
```

```
Db 644 IYKSOCKKKTTLDEFFENFKGLAKEKEMLEEAIDL-KALEKOKSWPSYTHVEKERA 702
Oy 93 EKLEKVPDYLEWYKAIGRTYIVDGDITKHSLSLEALSEDKKRIK-----IYKDAL 147
Db 703 STELSQ-----SRIKTSLYEIRSKTKRTASFIPTKESLT 738
Oy 148 HENHYAKEGEYEPVLIQSSPDYVENTEKALNVEYEGIKILSR-----DILSKINOPQK 202
Db 739 RDEQCCKEKRELQMRLEKE--ISHNENKMDFFSSKEGOYKAKIKELNNLERLSLDOS 796
Oy 203 FLVDLNTTKNASDS-----DGODLLFTNQKE-----HP 231
Db 797 KIOEESTIRSCDKQOLKRAQNTIDTEKMKMSLTLSNKTETTELKSTETENDKELRK 856
Oy 232 TDSVEEFLQNSNEVOEYFAKAFAYTEPQHR---DVLQYAPAFNMYDKNE--QGIN 286
Db 857 TKFOYKFLDQNSD-----ASTLEPTLRKELEQIOVOLKANQSOIQAYEETISSNE 906
Oy 287 LSLFEIKQRM-----LSRYEKWEKIKQHYQMSDSLSEEGGLKKTQIPIEP 335
Db 907 NALIEKNEIAKTKENYDAKTELEKKEKWAR-----EEDLSRL-RGELGEIR-ALQP 956
Oy 336 K-KDDIHSLSQEEK--ELKRIQ--IDSSDFLSTEKEFLKQIQRDLSSEEEK--- 387
Db 957 KIKGALHFVQOSERLNRVEERIQMKIEIKMST-IYQLCKKEMSGYSTMKENKDL 1015
Oy 388 ELNRIQVDSNPLESEKEFLKTKLDIQPYDINQRLDTGGLIDPSIN-LDVNRQYK 446
Db 1016 ELVIRLEKMDAQAE-----LTKTK-----SLYSADQLDKHERKWEKADYE 1061
Oy 447 RD-IQNTDALLHQSIGSLYKIKTYLENMNNINLTATIGALVSTDPKTKNRGIFNEFK 505
Db 1062 RELISNIEQT-----ESLRVEN-----SVLIRKVDYDTANNQDKHLK 1099
Oy 506 KNFYSSISNMYIVINERPALDNERLTKRIQSLPDRFAGYLENGKLLIQNRIGLE--IK 563
Db 1100 ---LVSLFSNLR-----HERNSLETKLTCKRELA-----FYKQKNDSELEKTIN 1140
Oy 564 DVQIILKQ-SEKEYIRIDAKVVPKSKIDTKIOEAOALNIOENKNAKGLPKYTKLITFNVHN 622
Db 1141 DLQRTQTLSEKEY-QCSAVIIDEFKDITK-EVTQVNIILKE--NNAI-LQKSLKNTYKRE 1196
Oy 623 RYA-----SMIV-----ESAYLLNEMKNNIOSD 646
Db 1197 IYKOLNDRQEIISRLQRLDIQTKQOVSTNSKNTILVESEMQCKQRYODLSQOKDAOKK 1256
Oy 647 LIKRYTNVLYVNGRFEVFTDITLPNIAEQYTHODEIYEQVHSHK----- 689
Db 1257 DIEKLTINISDLKGLKLSAEANANDLENKF--NRKKQAHEKIDASKQOQALTNELNE 1313
Oy 690 -----GLVPEERSIILHGSPKGVLELRNDSSEGFIEFGHAVDDYAGYLLDKNOSD 739
Db 1314 LKAIKDKLEQDLHEFNMAVVIDLDPKLAHNELOSEBVSRDH-----KDY-----RTLME 1363
Oy 740 LVYNSKKRPDIFKEEGSLTGYGRTNEAEFFAFRLMSTHDAERLKVKOKNAPKTFQFI 799
Db 1364 EIESIKRELQIFKYANS-----SSDAFEKLVKVNKEKRD-RII 1400
Oy 800 NDOIK 804
Db 1401 DERTK 1405

RESULT 11
ID MYSJ DICDI STANDARD; PRT; 2245 AA.
AC P54697;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN II HEAVY CHAIN.
GN MYOJ.
OS Dictyostelium discoideum (Slime mold).
```


DT 01-OCT-1996 (Rel. 34, last annotation update)
DE HYPOHETICAL 222.8 KDA PROTEIN C1F3.06C IN CHROMOSOME 1.
GN SPAC1F3.06C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
CC Schizosaccharomycetales; Schizosaccharomycetaceae;
CC Schizosaccharomycetes.
OK NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 270690; CAA94624.1; -
KW Hypothetical protein.
SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 5.1%; Score 212.5; DB 1; Length 1957;
Best Local Similarity 18.8%; Pred. No. 0.25;
Matches 188; Conservative 155; Mismatches 342; Indels 317; Gaps 44;

QY 59 FERNKTOEHLKLEIKHVIKVEKGEAVK-----EAAKLEKLEKPSDYLEMYKAIG-GK 113
DB 650 DKESTLSNDNDLTTLTKLEESKSLTKQEDVDSLEKNQOTLKEDLRKSEALRPSK 709
QY 114 I-----YIVDDITKHISLEALSED-KKIKIDYGDALHEHYVYAKEGEYEPV----- 161
DB 710 LEAKNLREVIODLKKKHETLEQNRDHLSSLDKNTNALISSELTSSSEYKRLTANVE 769
QY 162 -----LVYOSSEDEVVENTEKALNVEE----- 183
DB 770 TLTQDSKAMKQSFSTLVNSYQISNLYHELRDHYNMOSQNTLTSESSEKLTDCENTLTQ 829
QY 184 -----IGKILSRDI-----LSKINQYQKFLDYLVNTIKNASDSQDGLFTNTQL 227
DB 830 QNMTLIDNVQKLMHKKVQVESKSEVSELKSLDLKRLSRSLNVAISDNDQIL--TQL 887
QY 228 KEHPTDFSEVEFLEQNSNEVEQFAKAFAYVIEPQHRDYLQALVAPAFNYMDKF----- 280
DB 888 AELSKNY--DSLQESQALNSGLKSLA-----EKQLHTNEHELHTRDLKLTGKLTIE 939
QY 281 -----NEQET-NLSLEELKQORMLRYE-----KWEKIKQHYQHMSDSL 318
DB 940 ESKSSDLKKTLLARQEEISNLKEENMSQOAITSVKSKIDETLSKSKLEADIEHLKNVY 999
QY 319 SE--EGRGL-----KLIQPIEPK-----DQIIHSL----- 344
DB 1000 SEVEVERNALALASNERLMDLKNNGENIASLOTETIEKRAENDDQKSLSVSSEYENLL 1059
QY 345 ---SOEEREL-----LKRIO-----IDSSDFLSTEEKEFLKLT----- 374
DB 1060 LISSQTNKSLDEKTNQKLYIEKNVQKLLDEKQNRNVELEELSKGKIGEENAQIKDELL 1119
QY 375 -----QIDIRSLSEEEKELNLRIQVDSN--PLSEKEKEFLK 410
DB 1120 ALRKSKKQHDLCANFYVDLKEKSDALQTLTEKNEMLLYLSLEQSNNSNEALVEERSDLAN 1179
QY 411 KTKLDIQYDINQRLQDGGUL-----DSPSIN--LDVRK-----QYKDIQNIQAL 455
DB 1180 RLS-----DKKSLSDSDNNVSVIRSDLVARNDELDTLKKDKDSLSTQYSEVCCDRDRL 1233
QY 456 LHSQIG-STLVNK--IYLYENNNINNTATGADLVDSQDNTKIRGIFNEKK----- 506

DB 1234 LDSLKCESEFNKYNVAVSELCTKSEIDVPV--SEIID--DNFVNAGNFSLSRLTVLSL 1290
QY 507 -----NFKYSISSNYM-----IVDINERPALDNERLKWRIQSLPPTRAGY 546
DB 1291 ENYLDAPFNQVNRKKNKMLQRLTTTDAEFTKKVADLEK--LQHEHDDMLIQ-----RD 1341
QY 547 LENGKLIQRLNIGLEIKDQVQITKQSEKEYIRIDAKV-----VPSKIDTKIQEQALNIN 600
DB 1342 LE-----KALKDESEKNFLRKEAEWENTENIHSLERKEETKEKKEIAELSSR 1384
QY 601 QGMNVALGSLPKYTKLITTFVNHVRVANSIV--ESAVILINENKNN--IQSLDI--KKT 652
DB 1385 LBDNQATNKKLNQDLHQETRLKEDVLKERESESLTISLEESLSQROKESLLDQAKNEL 1444
QY 653 NYLVQNGRFFVTDTLPNIAEQYTH--ODEIYE-----QVHSKGLVYPSRS 698
DB 1445 EHMDDTSR---KNSLMKIESINSLLDKRFELASANEKGLAQKLSSELSMENIK 1501
QY 699 ILLHGPSKQVELRNDSGEFHEFGHAV---DDVAGYLDKQO--SDLVNKKRFTIDFK 752
DB 1502 SOLQEAKEKIQV---DESTIOELDHEITRASKNNYEGKLEMDKDSIIRDSLENIEQLNLLA 1558
QY 753 EEGSNLTSYGRTEAEFFAEAFRLMHSQDHAERLWQKNAPK 794
DB 1559 EKSNAVKRLSTKESSEFIQFNSRLADLEYHKQVSELSGRSK 1600

RESULT 13
MPL1_YEAST
ID MPL1_YEAST STANDARD: PRT: 1875 AA.
AC 002455:
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN-LIKE PROTEIN MLPI.
GN MLPI OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C.
RX MEDLINE=93247549; PubMed=8483450;
RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Gen. Genet. 237:359-369(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
RT new open reading frames.";
RL Yeast 9:1349-1354(1993).
RN [3]
RP FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
RP REPAIR.
CC
CC -I- SIMILARITY: SOME, TO THE TPR ONCogene.
CC -I- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L01992; AAA34783.1; -
DR EMBL; X73541; CAA51948.1; -
DR EMBL; Z28320; CAA82174.1; -
DR PIR; S38173; S38173.


```

Db 248 IEQINREIMNK-----KDEIKSYISEI-----KEYKDKCTEISNSKRG 286
Qy 404 KER-EPLAKKLIDIOPIYDINORLODGTGLDPSINLDVR--KOYKRDIONIALLLHOSI 460
Db 287 KDKIEPLEFEK-----PNE-----ESNSKNVINEINENINSEQYLKIDIE--DAKQAST 335
Qy 461 GSTLYNNKYLKENNNINNL--TATLGADVDSDNFKING--IFNEK-----505
Db 336 KVELFHK--HEPTTISIFKESEILG--VETKSOKKINAEDIMEIEHNSSEIOTOV 388
Qy 506 KNFYSISSNMIYDINERPALDNERLKWRIQLSPDTRFAGYLENGKILORN--IGLEI 562
Db 389 KGPEENLNK-----LNPHNYDNA-----EDELNDKST--NAKYLIEINLESVKHNL 434
Qy 563 KDOVITKO-SEKEYIR--IDAKVVPKS--KIDRKIOQAQININQEWKALGLPKYT 613
Db 435 SEINIKOGGEIKYSKAKDIOKIKATSENAETKLEKVDQSN-----479
Qy 614 KLIFVHNKRASNIVESAVILNEMKN-----NIOSDLIKVNTLVN-----657
Db 480 -----YVNLNOITTEERNLITE-KNRLNGIDSTITNIGALKESKNGNEIGLEKLE 531
Qy 658 --GNGREFVETDILPNI-----AEQYTHODEIVEOVHSGKGLYVESRSIL 700
Db 532 EIGNKRLKVDITKKSINSTVGNFSLEFNNEDLNQYDENKNKINDYENKMG-----581
Qy 701 LHGSKVELANDSEGTHERFGHAVVDYAGYLLDKNOSDLVTSNKKFIDI-FKEGSGNLT 759
Db 582 -----EYNEPEGSINKISENLRN-----ASENTSDY--NSAKTLRLAEQKEKVL- 625
Qy 760 SYGRNEAEFEAFRLHSHSTDAERLKYOKNAKPTQFOINDOK 804
Db 626 ----LNKEE--EANKYLIDYKVKVESFRFINKMESLDKINEMIK 663

RESULT 15
MYH6_MESAU STANDARD; PRT: 1939 AA.
ID MYH6_MESAU STANDARD; PRT: 1939 AA.
AC P13539; 060562;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DR MYOSIN HEAVY CHAIN, CARDIAC MUSCLE ALPHA ISOFORM (MYHC-ALPHA).
GN MYH6.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F1B; TISSUE=Liver;
RC MEDLINE=95115033; PubMed=7815459;
RA Wang R., Sole M.J., Cukerman E., Liew C.-C.;
RA "Characterization and nucleotide sequence of the cardiac alpha-myosin
RT heavy chain gene from Syrian hamster."
RT J. Mol. Cell. Cardiol. 26:1155-1165(1994).
RN [2]
RP SEQUENCE OF 1630-1939 FROM N.A.
RC MEDLINE=86205859; PubMed=3458174;
RA Liew C.-C., Jandreski M.A.;
RA "Construction and characterization of the alpha form of a cardiac
RT myosin heavy chain cDNA clone and its developmental expression in the
RT Syrian hamster."
RL Proc. Natl. Acad. Sci. U.S.A. 83:3175-3179(1986).
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

```

```

CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY (BY
CC SIMILARITY).
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IO DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L15351; AAB59701.1; -.
DR EMBL; M12995; AAA37081.1; -.
DR InterPro: IPR000048; IO.
DR InterPro: IPR002928; Myosin_head.
DR InterPro: IPR001609; myosin_tail.
DR Pfam; PF00612; IO; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRODOM; PD000355; myosin_head; 1.
DR SMART; SM00015; IO; 1.
DR SMART; SM00242; MISC; 1.
DR PROSITE; PS0096; IO; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family;
KW Calmodulin-binding.
FT DOMAIN 1 782 MYOSIN HEAD-LIKE.
FT DOMAIN 783 812 IO.
FT DOMAIN 842 1939 COILED COIL (POTENTIAL).
FT DOMAIN 1439 1443 ATP-ALA.
FT NP_BIND 178 185 POLY(POTENTIAL).
FT DOMAIN 657 679 ACTIN-BINDING.
FT DOMAIN 759 773 ACTIN-BINDING.
FT MOD_RES 129 129 METHYLATION (TRT-) (POTENTIAL).
FT MOD_RES 697 697 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 707 707 ALKYLATION (SH-2) (POTENTIAL).
FT CONFLICT 1633 1633 O -> L (IN REF. 2).
FT CONFLICT 1651 1651 H -> Q (IN REF. 2).
FT CONFLICT 1686 1687 H -> DV (IN REF. 2).
FT CONFLICT 1693 1693 V -> G (IN REF. 2).
FT CONFLICT 1844 1844 K -> R (IN REF. 2).
FT CONFLICT 1879 1879 A -> T (IN REF. 2).
FT CONFLICT 1885 1885 E -> Q (IN REF. 2).
FT CONFLICT 1907 1907 E -> V (IN REF. 2).
FT CONFLICT 1928 1928 D -> N (IN REF. 2).
FT CONFLICT 1933 1935 OKM -> KR (IN REF. 2).
SQ SEQUENCE 1939 AA; 223626 MW; DBC8297DFE83115A CRC64.

Query Match 5.1%; Score 210; DB 1; Length 1939;
Best Local Similarity 18.6%; Pred. No. 0.31;
Matches 159; Conservative 159; Mismatches 261; Indels 256; Gaps 36;

Qy 29 PLVQAGGHDVGNHVKKEKNKDNKRKTOEHLKEIM--KHIVKVGGEA 86
Db 840 PLKSAFTEKEMA-NMKEEPRVAKESLEKSEARKKLEKKVSLLOEKNDLPQVQABD 898
Qy 87 VKKDAEKLEKVDVLEMKATIGKTIYDQGTTHISLEALSEDKKTKIDYGDAL 146
Db 899 NLNDAER-----CDOLIKKKIQLEA-----KVE-----923
Qy 147 LHEHYVAKGEYEVLVIOSSQEDVVENTEKALNYVEIGKILSDILSKINPOKFLDV 206

```

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 2, 2001, 13:49:05 : Search time 1166.37 seconds
(without alignments)
22387.614 Million cell updates/sec

Title: US-09-747-521-1
2430
Sequence: 1 atgaataataaaaaagaatt.....agttcattatcatactataa 2430

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estfun:*
2: em_estlin:*
3: em_estlhm:*
4: em_estlhm:*
5: em_estlhm:*
6: em_estlhm:*
7: em_estlhm:*
8: em_estlhm:*
9: em_estlhm:*
10: em_estlhm:*
11: em_estlhm:*
12: em_estlhm:*
13: em_estlhm:*
14: em_estlhm:*
15: em_estlhm:*
16: em_estlhm:*
17: em_estlhm:*
18: em_estlhm:*
19: em_estlhm:*
20: em_estlhm:*
21: em_estlhm:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the change being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	103.2	4.2	1337	5	BF630719 HVSME001
C 2	102.6	4.2	1025	13	CNS014J2
C 3	93.8	3.9	1101	13	CNS0021J
C 4	92	3.8	1059	13	CNS0022B
C 5	91.8	3.8	1036	13	CNS031MT
C 6	91.6	3.8	1036	13	CNS00599
C 7	90.8	3.7	1885	10	BE420745
C 8	89.8	3.7	1101	13	CNS005EV
C 9	89.2	3.7	1135	13	CNS033GO
C 10	88	3.6	1001	13	CNS0155H
C 11	87.4	3.6	576	13	CNS035N7
C 12	87.4	3.6	1101	13	CNS0039C

C 13	86.8	3.6	838	5	BF629227	BF629227 HVSME001
C 14	86.8	3.6	963	13	CNS0044L	AL054918 Drosophila
C 15	86.6	3.6	1020	13	CNS038G9	AL232578 Tetradon
C 16	86.6	3.6	1151	11	BG309087	BG309087 HVSME000
C 17	86.4	3.6	908	13	A2548467	A2548467 ENTER30TR
C 18	86.4	3.6	1101	13	CNS000K5	AL077453 Drosophila
C 19	86	3.5	520	13	CNS02LKT	AL202924 Tetradon
C 20	86	3.5	1223	13	B12981	B12981 T24D11-Sp6
C 21	85.8	3.5	879	13	CNS017RG	AL147405 Anopheles
C 22	85.8	3.5	982	13	AQ325799	AQ325799 nbx00021B
C 23	85.6	3.5	1101	13	CNS000SX	AL050813 Drosophila
C 24	85.4	3.5	1101	13	CNS0153V	AL104965 Drosophila
C 25	84.8	3.5	471	10	AI404883	AI404883 GH24733.5
C 26	84.8	3.5	1101	13	CNS016OP	AL107011 Drosophila
C 27	84.6	3.5	834	13	B12387	B12387 F21E20-Sp6
C 28	84.4	3.5	542	13	CNS002YA	AL098284 Drosophila
C 29	84.4	3.5	865	13	AQ324474	AQ324474 m9x00018B
C 30	84.2	3.5	998	13	CNS000YXQ	AL096968 Drosophila
C 31	84	3.5	819	11	BG369277	BG369277 HVSME1002
C 32	83.6	3.4	784	10	AL525973	AL525973 AL525973
C 33	83.6	3.4	907	13	CNS021J4	AL176953 Tetradon
C 34	83.6	3.4	942	13	CNS018GS	AL109318 Drosophila
C 35	83.2	3.4	717	13	CNS062NC	AL422542 T7 end of
C 36	83	3.4	870	13	AQ330286	AQ330286 nbx0046J
C 37	83	3.4	871	13	A2671726	A2671726 ENTHP05TF
C 38	82.8	3.4	843	13	A2551618	A2551618 ENTPV54TR
C 39	82.8	3.4	974	13	CNS001TT	AL075432 Drosophila
C 40	82.8	3.4	1147	13	B13042	B13042 T30M24-Sp6
C 41	82.6	3.4	849	13	A2546009	A2546009 ENTFW53TF
C 42	82.6	3.4	892	13	A2193174	A2193174 SP_1022_B
C 43	82.6	3.4	990	13	CNS006OI	AL065624 Drosophila
C 44	82.2	3.4	796	10	AL524807	AL524807 AL524807
C 45	82.2	3.4	835	13	A2188415	A2188415 SP_1011_B

ALIGNMENTS

RESULT 1	BF630719/C	standard; RNA; EST; 1337 BP.
ID	BF630719	
XX	BF630719;	
AC	XX	
SV	BF630719.1	
XX	XX	
DT	21-DEC-2000 (Rel. 66, Created)	
DT	21-DEC-2000 (Rel. 66, Last updated, Version 1)	
XX	XX	
DE	HVSME0013H16f Hordeum vulgare seedling shoot EST library HVCNMA0002	
DE	(Dehydration stress) Hordeum vulgare cDNA clone HVSME0013H16f, mRNA	
DE	sequence.	
XX	XX	
KM	EST.	
XX	XX	
OS	Hordeum vulgare (barley)	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;	
OC	Triticeae; Hordeum.	
XX	XX	
RN	[1]	
RP	1-1337	
RA	wing R., Close T.J., Kleinhofs A., Wise R., Begum D., Frisch D., Yu Y.,	
RA	Anderson H., Dale J., Henry D., Kernodle S., Palmer M., Rambo T.,	
RA	Saski C., Schwartzbeck J., Simmons J., Choi D.W., Main D., Wood T.,	
RT	"Development of a genetically and physically anchored EST resource for	
RT	barley genomics";	
RL	unpublished.	
XX	XX	
CC	Contact: Wing RA	
CC	Clemson University Genomics Institute	
CC	Clemson University	
CC	100 Jordan Hall, Clemson, SC 29634, USA	

CC	Tel: 864 656 7288
CC	Fax: 864 656 4293
CC	Email: r.wing@clemons.edu
CC	Seq primer: ATTATACCCCTACTATAGGC
CC	High quality sequence start: 38
CC	High quality sequence stop: 1204.
XX	
PH	Key
PH	Location/Qualifiers
FT	source
FT	1. .1337
FT	/db_xref="taxon:4513"
FT	/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: Xho1"
FT	/organism="Hordeum vulgare"
FT	/cultivar="Morex"
FT	/clone="HVSMEB0013H16f"
FT	/clone_lib="Hordeum vulgare seedling shoot EST library
FT	HVCDNA0002 (Dehydration stress)"
FT	/tissue.type="Seedling shoot"
FT	/lab_host="JVC121"
XX	
SO	Sequence 1337 BP; 292 A; 22 C; 74 G; 941 T; 8 other;

Query Match	4.28;	Score 103.2;	DB 5;	Length 1337;
Best Local Similarity	46.18;	Pred. No. 3.2e+06;		
Matches 530;	Conservative	0;	Mismatches 602;	Indels 18;
				Gaps 5

[illegible][illegible]

```

RESULT 2
CNS014J2/c
LOCUS
DEFINITION
CNS014J2 1025 bp DNA GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN11L1 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL104216
ACCESSION
AL104216.1 GI:5615827
VERSION
KEYWORDS
SOURCE
ORGANISM
Plasmid Drosophila melanogaster
Eukaryote; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1025)
Genoscope.
AUTHORS
Direct Submission
TITLE
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
COMMENT:
- Web : www.genoscope.cns.fr)
- Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (DrosBAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
Project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES
Source
Location/Qualifiers
1..1025
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN11L1"
/note="end : T7"
BASE COUNT
148 a 88 c 6 g 522 t 261 others
ORIGIN

```


Db 499 AAAAAAAAAAATAATTAATTTTWTATTAATTAATTTTTTTTTTTTTTTTTTT 558
QY 494 aatctcggaagatatagtagaataaactgaaagcactgaacgtttatatatgaataag 553
Db 559 WTTAATTTTAAATTTTAAWMAAAATTTTAAATAAAWMTWTWTAAATTTAAWMAA 618
QY 554 gtaagatatctcaagagatatattgaataaatcaatcaacatatcagaattcttag 613
Db 619 AAAAAWTTTAAAWMTWTWTWTAAATTAATTAATAWMAAAWMTWTWTAT 678
QY 614 atgatatcaatcaatcaatcaatcgtga--ttcagatgagcaagatctttattac 671
Db 679 WMAATTAATTAATTTTAAWMTWTWTAAATTTTAAATTTTAAATTTTAAATTTTAA 738
QY 672 taatcagcttaaggaacacccacagactttctgtagaattctggaacaatagcaa 731
Db 739 AAAAATWMTWTWTAAATTTTAAWMTWTAAWMAAAATTTTAAATTTTAAATTTT 798
QY 732 tgaagtcacaagaagtattctggaagccttgcataatatcagccagcagctgta 791
Db 799 TTAATTTAAWMAAAWMTWTWTAAATTTAAWMAAAWMTWTWT 858
QY 792 tgtttacagcttatagcaccggaagctttaattacatgataaatttaacggaacaa 851
Db 859 TAAATTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 918
QY 852 aataatcatctcttggagaactaaagatcaacgagctgctgaagatggaataatg 911
Db 919 AWTGGAATGKAKKGGTAAARAATDGTAAAGAKRAATKATKATKTA 978
QY 912 ggaagaagataaagcctatcaacactgagcagatcttattctgaagaagaagg 971
Db 979 AAAAAAKKGAATAATGAARGAATGAAAAAGTGAATTTTAKKAAWMAAAWMAA 1038
QY 972 actttaaaaaagctgcagat 992
Db 1039 AAAAAAAAAAAGAGAKAK 1059

RESULT 4
LOCUS CNS0022B/c 1059 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN01115 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL097133.1 GI:5608744
VERSION GSS.
KEYWORDS fruit fly.
SOURCE plasmid Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Braachyera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1059)
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaut at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES
source
1. 1059
Location/Qualifiers
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"

Query Match 3.8%; Score 92; DB 13; Length 1059;
Best Local Similarity 37.5%; Pred. No. 0.0015;
Matches 348; Conservative 87; Mismatches 491; Indels 3; Gaps 1;
BASE COUNT 14 a 21 c 66 g 760 t 198 others
ORIGIN
/clone_11b="DrosBAC"
/clone="BACN01115"
/note="end : T7"
QY 1093 tctcagaagaagaagagcttctaaagaatatacaatctgtagtagatctttact 1092
Db 972 KCTCTWCAABAACACWCBCKAGCHCYBCKBCKWMAAADRADCKCYKBABBAAD 913
QY 1093 actgagagaagaaggtttttaaagaactacaatatgatatcgtgctttacttga 1152
Db 912 AAAAAAACAACAGNMAAACAACAGNMAAACAACAGNMAAACAACAGNMAAACAACAGNMAA 853
QY 1153 gaagaagaagcctttaaataagatacaggtgtagtagtaatccttactcga 1212
Db 852 AAAAAAACAACAGNMAAACAACAGNMAAACAACAGNMAAACAACAGNMAAACAACAGNMAA 793
QY 1213 gaagaaggtttttaaagaactgaactgatatataccatatgatatca---a 1269
Db 792 AMACBCKACACBCKAC 733
QY 1270 aggttgcaagatacagaagggttaattgtagtcgcgtcaattcctgtaagaaga 1329
Db 732 AACSCAKABAC 673
QY 1330 cagatataaaggagatatacaaatatgtagtcttattacatcaatcattggaagtacc 1389
Db 672 AAAAAAADAADGAKAKAAATDKKKKKAAATGRDKKKADDAKAKAAAAA 613
QY 1390 ttgtacaataaattatttattgataaataatgataatcaataccttaacgcaacctta 1449
Db 612 DRGAHA 553
QY 1450 ggtcggaattgattgattccactgataataactaaataagtagaggtatttcaatga 1509
Db 552 AAAAAAHA 493
QY 1510 ttcaaaaaaatctcaatatagtagtattctgtactatgatgtgtgataaagaa 1569
Db 492 AAAAAAHA 433
QY 1570 aggcctgcatgataatgagcggttggaaatggaatccaatatataccagatactga 1629
Db 432 AAAAAAHA 373
QY 1630 gcagatatattagaataatggaagcttataatcaagaagaacatcgcttgaataaag 1689
Db 372 AAAAAAHA 313
QY 1690 gatgtacaataatgaagcaatccgaagaagaatatataagattgtagcgaagtagtg 1749
Db 312 AAAAAAHA 253
QY 1750 ccaagaagtaaatagatgatacaaaaatccagaagcagcggttaaatataatccagaa 1809
Db 252 AAAAAAHA 193
QY 1810 aataagaatagggttaccaaaatatacaagaagcttattacatcaagctgataataga 1869
Db 192 AAAAAAHA 133
QY 1870 tatgataccaatattgtagaagtgcttatttaatatgaaatgaatgaatgaatgaatga 1929
Db 132 AAAAAAHA 73
QY 1930 caaagtatctataaaaaaggttaacaa 1958

Db 72 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 44

RESULT 5

CNS03LMJ 1036 bp DNA GSS 17-MAY-2000

LOCUS Tetraodon nigroviridis genome survey sequence PUC-Or1 end of clone

DEFINITION 036N11 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION AL250012

VERSION AL250012.1 GI:7971024

KEYWORDS GSS: genome survey sequence.

SOURCE Tetraodon nigroviridis.

ORGANISM Tetraodon nigroviridis

REFERENCE 1 (bases 1 to 1036)

AUTHORS Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C., Weissenbach, J., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissbach, J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1036)

AUTHORS Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F., Saurin, W. and Weissenbach, J.

TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 1036)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES

source

1. 1036

/organism="Tetraodon nigroviridis"

/db_xref="taxon:99883"

/clone="036N11"

/clone_1lb="G"

/note="Genoscope sequence ID : COBG036CG06SP1-end : PUC-Or1"

BASE COUNT 1005 a 6 c 7 g 17 t 1 others

ORIGIN

Query Match 3.8%; Score 91.8; DB 13; Length 1036;

Best Local Similarity 43.4%; Pred. No. 0.00016;

Matches 420; Conservative 0; Mismatches 547; Indels 0; Gaps 0;

Db 270 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 329

Qy 1292 taattgtagcgcgaattcaattcttgatgtaagaagcagtaaaaggatattcaaa 1351

Db 330 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 389

Qy 1352 atattgtagcttatacatcaatccattggaagtagcttgatacaataattattgt 1411

Db 390 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 449

Qy 1412 atgaatatgatcatcaataacattacagcaacctagtgcgatttgatcca 1471

Db 450 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 509

Qy 1472 ctgtaatactcaaatcaatgaggtatttcaatgaattcaaaaaaattcaatata 1531

Db 510 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 569

Qy 1532 gtattctagtaactatatagttgatgataaagaagcccgcatagataatgagc 1591

Db 570 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 629

Qy 1592 gttgaatgagagatccaatcattacacagatcgcagcaggaattagaaatgaa 1651

Db 630 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 689

Qy 1652 agcttatatacaagaacatcggtctggaataaagatgltcaataatgaacat 1711

Db 690 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 749

Qy 1712 ccgaataaagaatataaagattgtagcgaagtagtcgaagaagtaaatagataca 1771

Db 750 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 809

Qy 1772 aaatcaagaagcacagttaaataataatcagaatggaataagcattaggtacc 1831

Db 810 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 869

Qy 1832 aatatcaagaactatcatcattcaagcgatagtagatgcatccatattgagaa 1891

Db 870 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 929

Qy 1892 gtgcttattataattgtagatgtagaagaataataatcatcaagtgatcttaaa 1951

Db 930 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 989

Qy 1952 taacaaa 1958

Db 990 AAAAAA 996

RESULT 6

CNS00599/c 1036 bp DNA GSS 03-JUN-1999

LOCUS Drosophila melanogaster genome survey sequence REF3 end of BAC #

DEFINITION BACR11111 of RCP1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL057797.1 GI:4932579

VERSION AL057797

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1036)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : segrete@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Oosawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1..1036
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR11116"
/note="end : TTT3"

BASE COUNT 64 a 56 c 41 g 701 t 174 others
ORIGIN

Query Match 3.8%; Score 91.6; DB 13; Length 1036;
Best Local Similarity 39.3%; Pred. No. 0.00017;
Matches 305; Conservative 65; Mismatches 405; Indels 1; Gaps 1;

Oy 1040 aagaaagagcttctaaagaatacaatgatagtagtcttcttctactcagag 1099
Db 797 aaaaaaaatccaaagacaaayaaaaaavcstctaataatagaaaaatrtccsms 738
Oy 1100 aaaaagagctttaaagaagctcaaatgatattcgtagtcttcttctgaagaaga 1159
Db 737 aaahaaactagaaaaaaaragavnaaaasghbaaktckcacaamaawtctaatt 678
Oy 1160 aagagctttaaagaatcagagtgtagtagtaactcttcttctgaagaagaaga 1219
Db 677 aabkaktctwaaaaaaavmrmrcacaaamaaaatgsvsttaacatrraaaaanrraa 618
Oy 1220 agcttttaaaagctgaactgatattcaaccatagatattcaaaaggttcgaag 1279
Db 617 aarvmaaaaaaaataatgkgaamagaaaaaaagactaataatavayaatgacamaa 558
Oy 1280 atacaggagagtgtagttagtcgcaatcaatcttgtagtgaagaagcgatataa 1339
Db 557 taaaakaaagtstraatatamkkaaaagaaanakaagrtddakkaaaaaaakaaatrtadaa 498
Oy 1340 gggatattcaaatatgtgctttatcatcaatccttggaagtaccttgtaataa 1399
Db 497 aaaaaaaakaaakaaatg-kbtttkbaaaaaaaagadammkagdgcgaaakmckaaawata 439
Oy 1400 aaattatttgtagtaaaatagaaatcaatcaactcaagaacccaggtgcgatt 1459
Db 438 aaatttaattgtatataaaamgatkkaaraaakaaaaaaananaaaaaaa 379
Oy 1460 tagtgattccctgataactaaatagaggtatttcaatgaatcaaaaaa 1519
Db 378 aaaaaaaananaaaaaaaananaaaaaaaananaaaaaaaananaaaaaa 319
Oy 1520 attccaatatagtagttagtaactatagtagttagtaataatgaagccgagc 1579
Db 318 aaaaaaaananaaaaaaaananaaaaaaaananaaaaaaaananaaaaaa 259
Oy 1580 tagaataagagcgttggaatcgaaatcacaatcaccagatctcgagcagagatc 1639
Db 258 aaaaaaaananaaaaaaaananaaaaaaaananaaaaaaaananaaaaaa 199
Oy 1640 tagaataagagagcttatatacaagaacatcgctgctggaataaagagtgtaaca 1699
Db 198 aaaaaaaananaaaaaaaananaaaaaaaananaaaaaaaananaaaaaa 139

Oy 1700 taattagcaatccgaagaagaataatagtagtgcgaagtagtgcgaagaagta 1759
Db 138 aaaaaaaananaaaaaaaananaaaaaaaananaaaaaaaananaaaaaa 79
Oy 1760 aatagatcaaaaatccagaagcagcttaataatcaatcggatggaataa 1815
Db 78 aaaaaaaananaaaaaaaananaaaaaaaananaaaaaaaananaaaaaa 23

RESULT 7
LOCUS BE420745
DEFINITION HMW002.B02 tTCC HMW Barley Leaf Library Hordeum vulgare cDNA clone
ACCESSION HMW002.B02, mRNA sequence.
VERSION BE420745
KEYWORDS BE420745.1 GI:9418588
SOURCE EST.
ORGANISM barley.
Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
1 (bases 1 to 1885)

REFERENCE
AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogilhare,Y.,
Pechioni,N., Quailset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
Sorrells,M., Warburton,M. and Wenzel,G.
International Triticeae EST Cooperative (ITREC): Production of
Expressed Sequence Tags for Species of the Triticeae
Unpublished (2000)
COMMENT
Contact: Hermann RG
Botanisches Institut der LMU
Menzinger Str. 67, D-80638 Muenchen GERMANY
Fax: 49 30 171683
Email: hermann@botanik.biologie.uni-muenchen.de
International Triticeae EST Cooperative (ITREC)
<http://wheat.pw.usda.gov/genome>.
Location/Qualifiers
1..1885

FEATURES
source

1..1885
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HMW002.B02"
/clone_lib="ITREC HMW Barley Leaf Library"
/tissue_type="leaf"
/dev_stage="14 day old"
/note="Vector: pBluescriptSK(-); 850 bp average insert
size."

BASE COUNT 1138 a 219 c 212 g 176 t 140 others
ORIGIN

Query Match 3.7%; Score 90.8; DB 10; Length 1885;
Best Local Similarity 38.8%; Pred. No. 0.00019;
Matches 386; Conservative 0; Mismatches 610; Indels 0; Gaps 0;

Oy 959 aagaagaagagctttaaagaagctgcagatccctatgagccaaagaagtgaca 1018
Db 890 aaaaaaaananaaaaaaaananaaaaaaaananaaaaaaaananaaaaaa 949
Oy 1019 taatcatcttattctcagaagaagaagctcttaaaagaaatcaaatgtagta 1078
Db 950 nnn 1009
Oy 1079 gtgatttttattctctgaggaagaagctttaaagaagctcaaatgtagtgcg 1138
Db 1010 aaaaaaaananaaaaaaaananaaaaaaaananaaaaaaaananaaaaaa 1069
Oy 1139 attcttattctgaagaagaagaagctttaaagaaatagaaatagtagtagtaac 1198
Db 1070 aaaaaaaananaaaaaaaananaaaaaaaananaaaaaaaananaaaaaa 1129

[illegible][illegible]


```

FH source 1. 838
FT /db_xref="taxon:4513"
FT /note="Vector: LambdaZAP; Site_1: EcoRI; Site_2: XhoI"
FT /organism="Hordeum vulgare"
FT /cultivar="Morex"
FT /clone="HVSMEB0010613f"
FT /clone_lib="Hordeum vulgare seedling shoot EST library"
FT HYCDNA0002 (Dehydration stress)
FT /issue_type="Seeding shoot"
FT /lab_host="TJC121"
XX
SO Sequence 838 BP; 268 A; 19 C; 58 G; 492 T; 1 other;

```

```

Query Match 3.6%; Score 86.8; DB 5; Length 838;
Best Local Similarity 47.8%; Pred. No. 0.00092;
Matches 357; Conservative 0; Mismatches 373; Indels 17; Gaps 3;

```

```

QY 1116 aaagctacaatgatctgcttcttctgaagaagaagagctttaagtag 1175
DB 789 AATTATAGATTATTAAGCCGTTATTTATCTGCTTAATATAAATTAATATAAAT 730
QY 1176 aatacaggtgagtagtaatccttcatcgaagaagaagagctttaaaagct 1235
DB 729 AAAATTAATTAATAAACATACCAATTAATATAAATAAATAATGATACATATAAT 670
QY 1236 gaacttgatataccaacatagatalttaacaaaggtcgaagatagagaggttaat 1295
DB 669 AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 610
QY 1296 -----tgatagcgcgtcaatccttgatgtaagaagaagcagtlataaaggagatctc 1348
DB 609 AATATATTTTAAACCTGATTAATTAATAAATAATTAATTAATTAATTAATTAATTAAT 550
QY 1349 aaaaatctgctcttattcatcaatccatctggaagctcgttacaataaattatc 1408
DB 549 TATATATATTTTAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 490
QY 1409 tgaatgaaataatgaatatacaataacccttcagcaagccaggtcgagattagtgatc 1468
DB 489 TATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 430
QY 1469 ccactgataactaaatataagagtagtttccaatgatacca--aaaaattcca 1526
DB 429 ATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 370
QY 1527 atatagtaattctgaatacatatagatctgtgataataatgaagagcctgcatlagaata 1586
DB 369 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 310
QY 1587 tggagcgttgaaatcgagaatccaattataccaagatactcgagcagagatattagaata 1646
DB 309 AAAAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 250
QY 1647 tggaa-----agcttatactacaagaagaacatcgctcgtggaataaagagctgtgcaaa 1698
DB 249 AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 190
QY 1699 ataatgaagaatccgaagaagaatataatagatctgtaagagagtagtgcgaagaagc 1758
DB 189 TTAATTAATTTTCATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 130
QY 1759 aaaaatgatacaaaaatcgaagaagcaggttaataatacaaggaatggaataaagca 1818
DB 129 AAAAAATATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 70
QY 1819 ctagggttaccacaatatacaagaagctt 1845
DB 69 TCATTTTAAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 43

```

RESULT 14

```

CNS00A4L
LOCUS CNS00A4L 963 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR20A24 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL054918
VERSION AL054918.1 GI:4935889
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 963)
REFERENCE
Genoscope.
Direct Submmission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Googawa and
Aaron Mammose in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
PI and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
Location/Qualifiers
source 1. 963
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR20A24"
/note="end : TET3"

```

```

BASE COUNT 504 a 25 c 41 g 128 t 265 others
ORIGIN

```

```

Query Match 3.6%; Score 86.8; DB 13; Length 963;
Best Local Similarity 41.4%; Pred. No. 0.00088;
Matches 208; Conservative 55; Mismatches 237; Indels 3; Gaps 1;

```

```

QY 132 aaaagagaagaagaataaagtgagaaatgaagaagaatgaagaagaataaagc 191
DB 344 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 403
QY 192 acagaagagcatttaagaagaatcatgaaacacatctgtaaaatagaagttaaaaggga 251
DB 404 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 463
QY 252 ggaagctgtctaaagaagcagcagaagaagcttctggaagaagctaccatctgtctt 311
DB 464 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 523
QY 312 agagaigtataagcaatcgaggaagatatatactgtgagtgtgatatgataaaga 371
DB 524 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 583
QY 372 tatactctagaagcattatctgagaataagaataaataaagaacattatggaaga 431
DB 584 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 643
QY 432 tgccttatacatga---acattatgatacgaaagaagatatgaacccgactctgtc 488
DB 644 WTATTTAKAAKAATTAATAAAMAAGAKAKTAAGAAAGDAAATAATADTKKKAATAAATAAATAAATAA 703

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 2, 2001, 13:49:05 ; Search time 54.73 Seconds
(without alignments)
10055.573 Million cell updates/sec

Title: US-09-747-521-1
Perfect score: 2430
Sequence: 1 atgaataataaaaaagaatt.....agttcattatctaactataa 2430

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/1na/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/1na/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/1na/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/1na/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2430	100.0	3291	1	US-08-021-601-1
2	2430	100.0	3291	1	US-08-082-849B-1
3	2430	100.0	3291	5	PCT-US94-01624-1
4	762.4	31.4	1368	1	US-08-021-601-5
5	762.4	31.4	1368	1	US-08-082-849B-5
6	762.4	31.4	1368	5	PCT-US94-01624-5
7	762.4	31.4	1425	1	US-08-021-601-7
8	762.4	31.4	1425	1	US-08-082-849B-7
9	762.4	31.4	1425	5	PCT-US94-01624-7
10	762.4	31.4	1524	1	US-08-021-601-9
11	762.4	31.4	1524	1	US-08-082-849B-9
12	762.4	31.4	1524	5	PCT-US94-01624-9
13	91.6	3.8	7218	1	US-08-232-463-14
14	66.4	2.7	5361	4	US-08-973-462-2
15	66.4	2.7	6152	4	US-08-973-462-1
16	66	2.7	4766	5	PCT-US93-07261-10
17	65	2.7	3095	6	5231168-1
18	65	2.7	19124	2	US-08-487-826B-13
19	64.2	2.6	8920	2	US-08-446-855A-1
20	64.2	2.6	8920	2	US-09-150-741-1
21	61.4	2.5	615	4	US-08-998-416-186
22	61.2	2.5	696	4	US-09-461-697-193
23	61.2	2.5	699	4	US-09-461-697-191
24	61.2	2.5	717	4	US-09-461-697-189
25	61.2	2.5	774	4	US-09-461-697-187
26	61.2	2.5	819	4	US-09-461-697-185
27	61.2	2.5	1669	4	US-09-461-697-184

28	59.2	2.4	2223	1	US-08-257-073-4	Sequence 4, Appl
29	58	2.4	5852	1	US-07-867-106-2	Sequence 2, Appl
30	57.2	2.4	5433	3	US-08-929-329-1	Sequence 1, Appl
31	57.2	2.4	19124	2	US-08-487-826B-13	Sequence 13, Appl
32	56.6	2.3	837	4	US-08-998-416-288	Sequence 288, App
33	56.2	2.3	5852	1	US-07-867-106-2	Sequence 2, Appl
34	55.4	2.3	6243	2	US-09-056-075-1	Sequence 1, Appl
35	53.6	2.2	636	4	US-08-998-416-1137	Sequence 1137, Ap
36	52.2	2.1	3337	1	US-08-072-610-1	Sequence 1, Appl
37	52.2	2.1	3337	2	US-08-719-822B-1	Sequence 1, Appl
38	52.2	2.1	3337	4	US-09-092-458-1	Sequence 1, Appl
39	51.8	2.1	658	4	US-08-998-416-595	Sequence 595, App
40	51	2.1	636	4	US-08-998-416-1137	Sequence 1137, Ap
41	50.6	2.1	1654	3	US-08-913-842-20	Sequence 20, Appl
42	50.6	2.1	1744	3	US-08-913-842-27	Sequence 27, Appl
43	50.6	2.1	1875	3	US-08-913-842-1	Sequence 1, Appl
44	50.6	2.1	1965	3	US-08-913-842-2	Sequence 2, Appl
45	50.4	2.1	663	4	US-08-998-416-191	Sequence 191, App

ALIGNMENTS

RESULT 1
US-08-021-601-1

; Sequence 1, Application US/08021601

; Patent No. 5591631

; GENERAL INFORMATION:

; APPLICANT: Leppla, Stephen H.

; APPLICANT: Klimpel, Kurt R.

; APPLICANT: Nichols, Peter J.

; APPLICANT: Arora, Naveen

; APPLICANT: Singh, Yogendra

; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Needle & Rosenberg, P.C.

; STREET: 133 Carnegie Way, Suite 400

; CITY: Atlanta

; STATE: Georgia

; COUNTRY: USA

; ZIP: 30303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/021,601

; FILING DATE: 19930212

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Spratt, Gwendolyn D.

; REGISTRATION NUMBER: 36,016

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 404/688-0770

; TELEFAX: 404/688-9880

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3291 base pairs

; TYPE: NUCLEIC ACID

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; ORIGINAL SOURCE:

; ORGANISM: Bacillus anthracis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 580..2907

; US-08-021-601-1

Query Match 100.0%; Score 2430; DB 1; Length 3291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 atgaataaaaaagaattatataaagtaattagtatgtcaatgtttaagtaacagcaatt 60
|||||
481 ATGAAATTAATAAAAAAGATTATTAAGTAATGTATGTCATGTTAGTAACGCAATT 540
QY 61 accttgatggctccgcgtcttacccttgtaacaggggcgggcggtlcatgtgatgta 120
|||||
541 ACTTGTAGTGTCCTCCGCTTTATCCCTTGTAACAGGGGGGGGGCTCATGGTATGTA 600
QY 121 ggtatgcagcttaaaagagaagaaataaagaatgaataagaaagaaagaaatga 180
|||||
601 GGTATGACGCTAAAAAGAAAGAAAAATTAAGATAGATAAGAAAAAGAAAGAA 660
QY 181 cgaataaacaacagagaagagcatttaaggaaatcgaacacatgttaaaaaatga 240
|||||
661 CGAATTAACACAGAGAGACATTTAAGGAAATCATGAACACATTGTAAAAATGAA 720
QY 241 gtaaaaggaggagaagcgtgttaaaaaaaggcagcagaagaagctacttgaagaatca 300
|||||
721 GTAAAGGGGAGAGAGCTGTAAAAAAGAGCAGCAAAAAAGTACTGAGAAAGTACA 780
QY 301 tctgattcttagagatgtataaagcaattggagaagaagatatattctgtgatgac 360
|||||
781 TCTGATGTTTGAAGATGTAAAGCAATTGGAGAAAGATATATTGTCGATGCTGAT 840
QY 361 attaacaacatatcatctttagaagcatlctgaagataagaanaaaataaagaacatt 420
|||||
841 ATTACAAAACATATATCTTTTGAAGCATTTATCTGAAGTAAAGAAAAATTAAGACATT 900
QY 421 tatggagaagaatgctcttactacatgaacatgatgtatatacgaagaagatatgaacc 480
|||||
901 TATGGAAAGATGCTTTATTACATGAACATTATGATATGCAAAAAAGAGATATGAACCC 960
QY 481 gtaactgtaatccaaactctcggaagatgatgtagaanaaaactgtgaagcactgaagctt 540
|||||
961 GTACTTGTAAATCCAAATCTTCGGAAGATTATGTAAAAATCATGAAAGGCACTGAACGTT 1020
QY 541 tatcatgaataaggtaagatatltaagaagatatlttaagtaaatltaacaacatat 600
|||||
1021 TATATGAATAGGTAAAGATATTTCAAGGATATTTTAAGTAAATAATTAATCAACCATAT 1080
QY 601 cagaatctttttagatgtataataataacataaanaaagcatctgaatcgaatgaacagat 660
|||||
1081 CAGAAATTTTATGATGTATTAAATACCATTAATAATGATCATGATCGATGACAGAT 1140
QY 661 ctttattactaaatcagcttaagaacatccccagaactttctgtagaattcttgaa 720
|||||
1141 CTTTATTTCTAAATCAATTCGAAACATCCACAGACTTTTGTAGAAATTTTGGAA 1200
QY 721 caaatatgacatgggtacagaagaatttgcgaagctttgcatactatcatcgagcca 780
|||||
1201 CAAATATGCAATGATGATCAAGAAGTATTTGCGAAAGCTTTTGCATATTAATGAGACCA 1260
QY 781 cagcatcgtgatgttcttaccagcttatgcacgggaagctttaatcatatgatataatc 840
|||||
1261 CAGCATGCTGATGTTTACAGCTTTATGACCGGAGAGCTTTAATTACATGATTAATAATTT 1320
QY 841 aacgaaacagaataaactctactctgtggaagaacttaagaatcaaggaatgcttcaaga 900
|||||
1321 AACCAAAAGAAATTAATCTATCTTGGAAGAACTTAAGATCAACGATCTCTCAGA 1380
QY 901 tatgaanaatggggaagaatgaanaacagcatatacaacactggagcggattcttcttga 960
|||||
1381 TATGAAAAATGAGAAAAATTAACACGACATATCAACCTGAGCGATTCTTTATCTGAA 1440
QY 961 gaaggaaagagacttttaaaaaagctgcagatctcacttgcgaagaagaagatgacata 1020
|||||
1441 GAAGGAAGAGACTTTTAAAAAAGCTGCAGATTCCTATTGAGCAAAAGAAAGATGACATA 1500

1021 attcaatcttattctcaagaagaagaagccttcaaaaaagaataacaaattgatatgac 1080
|||||
1501 ATTCAATCTTTATCTCAAGAAAGAAAAAGACCTCTAAAAAGAAATCAAAATGATATGAT 1560
QY 1081 gattcttactacttgaggaagaagcttcttaaaaaagctcacaaattgatatcgatc 1140
|||||
1561 GATTTTATCTACTGAGAAAAAGACTTTTAAAAAGCTACAAATGATATTCGAT 1620
QY 1141 tcttactcgaagaagaagaagcctttaaataagaatacagtgataglaatcct 1200
|||||
1621 TCTTATCTCAAGCAAGAAAAAGACCTTTAAATGAAATACAGGGGATATGATTAATCCT 1680
QY 1201 tttactgaaaaagaagaagcttcttaaaaaagcttgaacttgatatcaacatatgac 1260
|||||
1681 TTATCTGAAAAAGAAAAAGATTTTAAAAAGCTGAACCTGTGATTTCAACCATATGAT 1740
QY 1261 attaacaaagttgcagaatcacagagggttaattatgatccgtcaactatcctgac 1320
|||||
1741 ATTATTAAGGTTGCAAGATACAGAGGGTTAATTATATAGTCCGTCAATTAATCTTGAT 1800
QY 1321 gtaagaagagataaagaaggatltcaaaaatltgatgcttattatataccaatccat 1380
|||||
1801 GTAAGAAGCAGTATAAAGGATATTCAAAATATGATGCTTTATTAACATCAATCCATT 1860
QY 1381 ggaagtaccttgtaacaataatttatgttatgaanaatatgaatatcaataccttaca 1440
|||||
1861 GGAAGTACCTTGTAACAATAATTTATTTGTATGAAAAATATGAATATCAATTAACCTTTACA 1500
QY 1441 gcaacctgaagttgcggaattgaattgataccactgaataactcaaaatlaagaagatc 1560
|||||
1921 GCAACCTAGTGGGATTTAGTTGATTCACATGATATATCTAATAATTAATAGAGTAT 1980
QY 1501 ttcaatgaattcaaaaaaaatlccaatalaaglatlctagtaactatgatgtgtgat 1560
|||||
1981 TTCAATGAATTCAAAAAAATTTCAAAATATAGATTTCTGTAACATTAATGATGTTGAT 2040
QY 1561 ataaatgaaggcctgcattagataagaagcgttltgaagaagggaatccaattatccca 1620
|||||
2041 ATAATTAAGAGCCTGATTAATGATTAATGAGAGGTTTGAAGTGAATTCACATTAATCCCA 2100
QY 1621 gatactcgacagagatatlttagaanaatggaagccttatatacaagaagaacatcgctcgt 1680
|||||
2101 GATACCTGACAGAGATTTTGAAGAAATGGAAGCCTTATTAACAAACAAACATCGCTCG 2160
QY 1681 gaataaaggatgtacaataaattgaagcaatcogaanaagaataataagatgtatgcg 1740
|||||
2161 GAATTAAGGATGTACAAATTAATTAACCAATCCAAAAAGATTAATTAAGATGATGCGC 2220
QY 1741 aaagtatgcgaagaagtaaatgatgatacaaaaatltaagaagaagcagttaaatataat 1800
|||||
2221 AAATGATGCCAAAGAGCTTAATAATGATTAACAAAATTCAGAGAGCACAGTTAAATTAAT 2280
QY 1801 cagaatcgaataaagaatcaggttaccaaaatacaaaagcttactatcaatccaagtg 1860
|||||
2281 CAGGAATGGAATAAAGATTATAGGTTACCAAAATATACAAAGCTTATTAATTAACGTG 2340
QY 1861 catatagatatgcatcccaatatgtgagaagtgcttaattatataatcgaatggaatgaaa 1920
|||||
2341 CATATATGATATGATCATATATGTAGAAAGTCTTATTTAATTAATGAATGAAATGAAA 2400
QY 1921 aataatltcaaaatgtatcttataaaaaaggtlaaacaattactgaattgtatgtaagga 1980
|||||
2401 AATATATTTCAAAAGTGAATCTTAATAAAGGTAAACAAATTAATTAATGATGTAATGGA 2460
QY 1981 agattgtcttcttaccgatatctccctcaataatagctgacaacaatatcacatcaagat 2040
|||||
2461 AGATTTGTTTTACGATATTAATCTCTCCCTTAATATAGCTGAACAAATTAACATCAAGAT 2500
QY 2041 gagaatagagcaagttcatlcaaaagggttatatgtlccogaatcccgltctataat 2100
|||||
2521 GAGATATATGAGCAAGTTCAATTCAAAAGGTTATATGTTCAGAAATCCGTTATTAATTA 2580
QY 2101 ctccatggaaccttcaaaaagtgtagaatlaagaaatgaatgtagagggtttattacaagaa 2160

```
|||||
Db 2581 CTCATGACCTTCAAAAGGTGTAGAAATTAGGAATGATAGTGGGTTTATACACGAA 2640
Qy 2161 ttggacatgcgttgatgatatctgcatatctatgataagaaccaatctgatta 2220
Db 2641 ttggacatgcgttgatgatatctgcatatctatgataagaaccaatctgattta 2700
Qy 2221 gtacaaatctcaaaaatcatctgatttttaagaagaagaagatcttaacttcg 2280
Db 2701 GTTACAAATTTCTAAAAATTCATGTATTTTAAAGGAAGAGGATATTAACTTCG 2760
Qy 2281 tatggagaacaatgatgaacggaatttttcgagaaccccttaagttaatcattcag 2340
Db 2761 TATGGAGAACAATGATAGCGGAATTTTTCGAGAACCTTTAGTTAATGATTCAG 2820
Qy 2341 gacacatgcgaacgtttaaaagttcaaaaaaatgctccgaaaaacttccaattat 2400
Db 2821 GACCAATGCTGAAGCTTTAAAGTTCAAAAAAATGCTCCGAAAACTTCCAAATTA 2880
Qy 2401 gatcagattaagttcatattatcaacataa 2430
Db 2881 GATCAGATTAACTTCATTATTAACTCATTA 2910

RESULT 2
US-08-082-849B-1
; Sequence 1, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Kilpel, Kurt R.
; APPLICANT: Arora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/082,849B
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/021,601
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-161-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3291 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
```

```
; ORGANISM: Bacillus anthracis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 580..2907
; OTHER INFORMATION: /product= "lethal factor"
US-08-082-849B-1

Query Match 100.0%; Score 2430; DB 1; Length 3291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgaatataaaaaagaattataaagaatattagatagtcagtctttagtaagcaacatt 60
Db 481 ATGAATATTAATAAAGCAATTTATTAAGTAATGATGATGATGATGATGATGATGAT 540
Qy 61 actttagtgcgtccgtcttatacccccctgtacaagggcgggcggtcattgtagtga 120
Db 541 ACTTTGATGTCGCCGCTTTATCCCTTGTAACAGGGGGCGGGCGGTGATGATGATG 600
Qy 121 ggtatgcagtaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 180
Db 601 GGTATGCAGCTAAAGAGAAAGAGAAATTAAGATGATGATGATGATGATGATGATGA 660
Qy 181 cgaataaacaacaggaagaagcatttaagaagaatcagaacacatttgaataataga 240
Db 661 CGAATATTAACAACAGAGAGCATTTTAAGGAATCATGAAACATTTGTAATAATTA 720
Qy 241 gtaaaaggagggaagcgtgttaaaaaaaggcagcagaagaagcacttgaagaagta 300
Db 721 GTAAGGGGAGGAGAGCTGTAAAAAAGAGCAGCAAGAAAGCTACTTGAGAAAGTACA 780
Qy 301 tctgatgttttagagatgataaagaacattggagaagaagatataatctgtgtagt 360
Db 781 TCTGATGTTTATAGATGATTAAGCAATTTGGAGGAAGATATATATGATGATGATG 840
Qy 361 attacaaaacatatctttagaagcattctcgaagaatagaagaagaagaagaagaac 420
Db 841 ATTACAAAACATATATCTTTAGAACCTTATCTGAGAGATTAAGAAATAATTAAGAC 900
Qy 421 tatggagaagaatgcttattatatacgaacattatgatatgtaaaagaagaataga 480
Db 901 TATGGGAAGATGCTTTATATACATGAACATTATGATATATCAAAAGAGATATGA 960
Qy 481 gtaacttgtaaccatcttcggaagaatatagtagaataatctgaaagcagcagcgt 540
Db 961 GTACTTGTAATCCAACTTCTGGAGATATGTGAAATCTGAAAGGCACTGAAGCTT 1020
Qy 541 tattatgaagaatagatattatatacaaggatatttaagtaaaatatacaacatat 600
Db 1021 TATTATGAANTAGTAAAGATATTATCAAGGATATTTTAAGTAAATTAATCAAC 1080
Qy 601 cagaatctttagatglatataaataccatataaaatgcattgatacagatagaaga 660
Db 1081 CAGAAATTTTACATGATTAATAATACATTAAAAATGCATCTGATGATGCAAGAT 1140
Qy 661 ctttattactaatacagcttaagaacatccacagacttctgtagaatctctggaa 720
Db 1141 CTTTATTTCTAATCAGCTTAAGGAACATCCACAACTTTCTGTAGAAATCTTGGA 1200
Qy 721 caaaatagaataggtacagaagaatatttggaagaaccttgcattatatacagagca 780
Db 1201 CAAATAGCATAGGATACAGAGATTTTGCAAAACCTTTTGCAATTTATTCGACCA 1260
Qy 781 cagcatcgtgtagtcttatacagcttatacgcggaagcctttaaattatatacaga 840
Db 1261 CAGCATCGTATGTTTATACAGCTTTATGCAACCGGAAGCTTTTAATTATCATGAT 1320
Qy 841 aacgaacaagaataaatactataccttggaagaacattaaagatacagagatgctga 900
Db 1321 AACGAACAAGAAATTAATCTATCTTGGAAGAACTTAATAAGATCAACGATGCTGTA 1380
Qy 901 tatgaaaaaatgggaaaaagataaacaagcactatacaaacacgagcgagattcttata 960
```

|||||
Db 1381 TATGAAAAATGGAAAGTTAAACAGCATATVCAACACTGGAGGATCTTTATCTGAA 1440
Qy 961 gaaggaagagagactttaaanaagctgcagatctctatctgaagcgaagaagatgacata 1020
Db 1441 GAAGGAAGAGAGACTTTTAAAAAGCTGCAGATCTCTATTGAGCCAAAGAAAGATACATA 1500
Qy 1021 atcatctctatctcaagaagaagaagagctctcaanaagaatacaaatgtagtagt 1080
Db 1501 ATTCAATCTTTATCTCAAGAAAGAAAAAGAGCTTTAAAAAGAAATCAATATGATAGT 1560
Qy 1081 gatttttatactactgaaggaagaagaaggtttttaaanaagctcaaatgtagtagt 1140
Db 1561 GATTTTATCTACTGAGAGAAAAAGAGTTTAAAAAGCTTCAAAATGATATTCGTAT 1620
Qy 1141 tcttatctgaagaagaagaagagctttaaataagatagcaagtgtagtagtagtcc 1200
Db 1621 TCTTTATCTGAAGAAAGAAAAAGAGCTTTTAAATAGATACAGTGGATAGTAGTATCT 1680
Qy 1201 ttactgaaaaaagaagaagagttttaaanaagctgaaactgatatcaaccataagat 1260
Db 1681 TTATCTGAAAAAGAAAAAGAGTTTAAAAAGCTGAAACTGATATTCACATATGAT 1740
Qy 1261 attaatcaaggtgcaagatatacaagaggttaattgtagtagcgctcaattatctgat 1320
Db 1741 ATTATTCAAAGGTTCCAAAGATACAGAGAGGTTTAAATGATAGTCCCTCATTTATCTGAT 1800
Qy 1321 gtaagaagaagatataaagaagatatacaaatatgtagtcttatacatcaatcatt 1380
Db 1801 GTAAAGAAAGCATATAAAGAGATATCAAAATATTTGATGCTTTATTCATCATATCAT 1860
Qy 1381 ggaagtagctgtagataaataattatctgtatgataaataatgaataatcaacttaca 1440
Db 1861 GGAATGACCTGTGTAATAATTTATTTGATGAAAAATGAAATATCAATTAACCTTTACA 1920
Qy 1441 gcaaccacagtgagcagattagtgatccactgataataactcaaatgaagagatt 1500
Db 1921 GCAACCCATAGGTGGCGATTTAGTTGATTCACACTGATATCTAAATTAATAGAGTATT 1980
Qy 1501 ttcaatgaatcaaaaaaattcaaatatgtagtcttctagtaactatagtttgat 1560
Db 1981 TTCAATGAAATCAAAAAAATTTCAAAATATGATATTTCTAGTAATATGATGTTGTAT 2040
Qy 1561 ataaatgaagaagcgtgctgattgataatagagcttgaagaatgagaatccaatatcaca 1620
Db 2041 ATAAATGAAAGGCTGCTGATATGATATGATAGAGCGTTTGAATGAGAAATCAATATACCA 2100
Qy 1621 gatactcgagcagagatcttgaanaatggaagcttatatacaagaagaacatcgctg 1680
Db 2101 GATACTCGAGCAGGATTTTGAATAATGGAAGCTTATATTTACAAAGAAACATCGGTCTG 2160
Qy 1681 gaataaagaagatgtacaataatlaagcaatccgaanaagaatatataagattgtag 1740
Db 2161 GAAATTAAGAGATGTACAAATATATTAAGCAATCCGAAAAAGATATATTAAGATGATCG 2220
Qy 1741 aaagtagagcgaagaagaatagatacaaaatctcaagaagcgcattttaaataat 1800
Db 2221 AAAAGTAGGCCAAAGAGATTAATAGATACAAATTTCAAGAGCAGCTTAATTAAT 2280
Qy 1801 caggaatggaataaagcctagaggttaccaaaataatacaagcttatcatcaacagtg 1860
Db 2281 CAGGAATGGAATTAAGCAATTTAGGCTTACCAAAATATACAAAGCTTATTAATCAACGTG 2340
Qy 1861 cataatagatgtacccaataatgttagaaggtcttatctaattgaaatggaatgga 1920
Db 2341 CATATATATATGATCCAAATATTTAGTAAAGTCTTATTTATTTGAATGAATGAGAA 2400
Qy 1921 aataatatacaaggtactataaaaaagtaaaaaagtaaaaaatctactgtgtagtga 1980
Db 2401 AATAATATTTCAAAGGATCTTATATAAAAAAGTTAAACAAATTTACTGTGATGATATGA 2460
Qy 1981 agattgttctacagatctactctcctataatagctgaaacaatatcacataaat 2040
|||||

Db 2461 AGATTGTTTTTACCGATATTACTCTCCCTAATATATAGCTGAACAAATATACATCAAGAT 2520
Qy 2041 gagatataagcaaggttcatctcaaaaggtatatagtctccagaaatcccgcttcaata 2100
Db 2521 GAGATATATGACAGAGTTTCATTCAAAAGGCTTATATGTTCCAGAAATCCCGTTCTATTA 2580
Qy 2101 ctcaatgagacctcaaaaggtgtagaattgaagaatgaatgtagtagtcttatacagaa 2160
Db 2581 CTCATGAGACCTTCAAAAGGCTAGAAATTAAGAAATGATAGTAGAGGTTTATACAGCA 2640
Qy 2161 tttagacatgctgtagatgatatgtctgatatctatagataagaacaactcgattta 2220
Db 2641 TTTGACATGCTGTGATGATATGCTGATATCTATTAAGTAAACCAATCTGATTTA 2700
Qy 2221 gttaacaattctaaanaatcaatgataattttaaagaagaagagagtaacttaactcg 2280
Db 2701 GTTCAAAATTTCAAAATTTTCATTGATTTTAAAGAAAGAGAGATTTAATTAACCTCG 2760
Qy 2281 tatggagaacaanaatggaagcgaatcttctgaagaagccttgaagttaaatgcatctag 2340
Db 2761 TATGGAGAACAAATGAAGCGGAATTTTTCAGAAAGCTTTAGGTTAATGCATCTCAG 2820
Qy 2341 gaccatgctgaagctttaaagaatcaaaaaaagctccgaanaacttcccaattatrac 2400
Db 2821 GACCATGCTGAACGTTTAAAGTTCAAAAAAATGCTCCGAAACTTTCCAAATTTATTAAC 2880
Qy 2401 gacagattgaattcatatataactacataa 2430
Db 2881 GATCAGATTAAAGTTCATTATTAACATCAATTA 2910

RESULT 3
PCT-US94-01624-1
Sequence 1, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Lepola, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Alora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESS: TOWNSEND and TOWNSEND KHOURIE and CREW
STREET: Stewart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3291 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: CDS
LOCATION: 580..2907
OTHER INFORMATION: /product= "Lethal Factor"
PCT-US94-01624-1

Query Match 100.0%; Score 2430; DB 5; Length 3291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgatatataaaagaattatataaagttaattgatatgcatgctttagtaacgaatt 60
DB 481 ATGATAATATAAAAGAAATTTATAAAGTAAATGATGTCATGTTTAAACAGCAATT 540
QY 61 accttgatggtccgcgtcttatcccccgttacaggggggcggtcgaatgcatgta 120
DB 541 ACTTTGATGCTGCCGCTTTATCCCTTGTAAGGGGGGGCGGCTCATGTGATGTA 600
QY 121 ggtatgcagtaaaagaagaataaataaagatgagaataaagaagaagaaga 180
DB 601 GGTATGCACGTAAAGAGAAAGAGAAAATTAAGATGGAATAAGAGAAAAGATGAAGA 660
QY 181 cgaataataaacacaggaagagcatttaagaataatcagaaacacattgtataaata 240
DB 661 CGAAATATAACACAGAGAGCATTTAAAGCAAAATCATGAACCATTTGTAATAATGAA 720
QY 241 gtaaaaggaggaagagcgtttaaaaaagagcagcagaagaagcacttgagaaga 300
DB 721 GTAAGAGGAGAGAGCTGTTAAAAAGAGCAGACAGAAAACACTTGACAAAGTACCA 780
QY 301 tctgaatctttagagatgataaagaattggagaagaatataattgtgagtgat 360
DB 781 TCTGATGTTTATAGATGTTAAAGCAATTGAGAGAAAGATATATTTGTCGATGTCAT 840
QY 361 attacaaacatatatctttaagaagcattatctgaagataagaataaataaagaacat 420
DB 841 ATTACAAACATATATCTTTTGAAGCAATTTCTGAAGATGAAGAAAATAAAGAAATTT 900
QY 421 tatggagaagatgcttattacatgaacatgatatgatacaaaagaagatagaacc 480
DB 901 TATGGGAAGATGCTTTATATACATGAACATTTATGATGCAAAAAGATGTAAGACCC 960
QY 481 gtaactgttaacaaatcttcggaagattatgtagaataactgaaaagcactgaacgt 540
DB 961 GTAATTGTATCCAAATCTTCGGAAGATTATGTGAATAATCTGAAAAGGCACCTAACGTT 1020
QY 541 tattatgaataatgtaagaatatatcaaggagataatttaagtaaaatcaacat 600
DB 1021 TATTATGAATAATAGTAAAGATTTATCAAGGATATTTAAGTAAATAATATCAACCATAT 1080
QY 601 cagaataatttagatgattataaatacctaataaattgcattgatttaagatgagaagat 660
DB 1081 CAGAAATTTTATGATGATTAATACCATTAATAAATGCAATCTGATTCAGATGGCAAGAT 1140
QY 661 ctttatttactaatcagcttaaggaacatccacagacttctctgagaattcttgaa 720
DB 1141 CTTTATTTACTATATGAGCTTAAGGAACATCCACACACTTTTCTGTAGATTCCTTGGA 1200
QY 721 caaaatagcaatgaggtacagaagatcttgagaagccttgcatattatlatcgaca 780
DB 1201 CAAAATAGCAATGAGGTACAGAAAGATTTGCGAAACCTTTGATATTTATTCAGACCA 1260
QY 781 cagcatcgatgcttttaagagctttatgacacggaagcctttaaatacatcggataaatt 840
DB 1261 CAGCATCGTATGTTTATACGCTTTATGACCGGAGACCTTTTAAATTAATCATGAGATAATTT 1320
QY 841 aacgacaagaataataatctatctcttgagaagacttaagaactcaagatgctgtaaga 900

DB 1321 AACGAACAGAAATAAATCTATCCCTGGAGACACTTAAAGATCAACGATCTGTCAAGA 1380
QY 901 tatgaaaaatgagaaagatataaacagcactataaacctggagcgattcttctcgaa 960
DB 1381 TATGAAAAATGGGAAAGATTAAGAACACATATCAACCTGAGACGATTTCTTTATCTGAA 1440
QY 961 gaaggaagagacttttaaaaagctgcagatttcttacttgagccaaagaagaatgacata 1020
DB 1441 GAAGGAAGAGACTTTTAAAAAAGCTGCAGATTCCTATTCAGCCAAAGAAAGATGACATA 1500
QY 1021 attcattcttatctcaagaagaagaagcttcttaaaaagctgaactgtatctcaaccatgat 1080
DB 1501 ATTCAATCTTATATCTCAAGAGAAAGAGCTTCTTAAAAACATATCAATGATAGTAGT 1560
QY 1081 gatttttactactgaggaagaagagcttttaaaaagctgacaaattgatattcgtgat 1140
DB 1561 GATTTTATCTACTGAGAAAGAGCTTTTAAAAAGCTACAAATTGATATTCGTGAT 1620
QY 1141 tcttattctgaagaagaagaagcctttaataatgaatacaggtggaatgaatccct 1200
DB 1621 TCTTATCTGAGAGAGAAAGAGCTTTTAAATGAAATACAGGTGATGATGATTAATCCT 1680
QY 1201 ttatctgaaaaagaagaagatttcttaaaaagctgaactgtatctcaaccatgat 1260
DB 1681 TTATCTGAAAAAGAAAGAGCTTTTAAAAAGCTGAAACTGATGATTCACACATATGAT 1740
QY 1261 attaatcaaaaggttgcagaatgacagaaggttaattgatatgctgcgaatlaactgat 1320
DB 1741 ATTATATCAAGGTTGCAAGATACAGAGGGTTAATGATGATCCGTCGAATTAATCTGAT 1800
QY 1321 gtaagaagaagcagtaataaagagatatcaaatatgattcttattacttaactcaatc 1380
DB 1801 GTTAAGAAAGCATATTAAGAGGATATTCAAAATTTATGCTTTTATACATCAATTCATT 1860
QY 1381 ggaagtaacctgtgataaataattatctgatatgaataatgaataatcaactca 1440
DB 1861 GGAAGTACCTGTGATCAATTAATTTATTTGATGATGATGATGATGATGATGATGATGAT 1920
QY 1441 gcaacctgaagctgcgagattagttgattccaacgataactcaataatlaagaagatt 1500
DB 1921 GCAACCTGAGTGGGATTTAGTATTCACATGATTAATCTAAATAATTAATAGAGTAT 1980
QY 1501 ttcaatgaattcaaaaaaatttcaaatatagatatctctgtaactatagattggtgat 1560
DB 1981 TTCAATGAATTCAAAAAATTTCAAAATATGATTTCTAGTAACTATATGATGATTTGAT 2040
QY 1561 ataaatgaagagcctgcatlagataatgagcgtttgaaatggaagaatccaattaccca 1620
DB 2041 ATTAATGAAGAGCCTGATTAAGATTAAGACGTTTGAATGAGCAATCCATTTATCCCA 2100
QY 1621 gatactcgacagagatattagaanaatggaagcttatattacaagaacatcggtctg 1680
DB 2101 GATTACTGACAGAGATATTTAGAAATGGAAGCTTATTTACAAAGAACATCGGCTG 2160
QY 1681 gaataaagaatgtacaaataatlaagcaatccgaaagaataatataaagattgagcg 1740
DB 2161 GAAATTAAGATGTACAAATTAATTAAGCAATCCGAAAAGAAATATTAAGATTGAGCGC 2220
QY 1741 aaagtaatgcaaaagataaataatagatacaaaaattcaagaagcaagtttaataataat 1800
DB 2221 AAAGTATGCGCAAAAGATTAATTAAGATTAAGCAAAATTTAAAGACACAGTTTAATTAAT 2280
QY 1801 caggaatggaataaagcattaggttaccaaaatatacaaaagcttatatacacaagtg 1860
DB 2281 CAGGAATGGAATTAAGCATTAAGGTTTACCAAAATATTCAAAAGCTTATTAATTAATTAAT 2340
QY 1861 cataatagatatgcatcccaatattgtagaagaagtgctatttaatatgaaatgaaatg 1920
DB 2341 CATTAATGATATCATCAATATTTAGAAAGCTTATTAATTAATTAATTAATTAATTAAT 2400
QY 1921 aataatataaaggtatctataaaaaggtiaacaattacttagttgltgtaatgga 1980

```
Db 2401 AATATATTCAGATGCTTATTAAGTACAAATTACTAGTGTGTAATGCA 2460
Qy 1981 agatttcttaccgatactctcccaataatagctgaacaatatacacatcaagat 2040
Db 2461 ACATTTGTTTACCGATATCTACCTCCCTAATATAGTGAACAATATACATCAAGAT 2520
Qy 2041 gagatataagacagatctacaaaggctatatagtctcagaatcccgcttatata 2100
Db 2521 GAGATATATGACCAAGTTCATTAAGGGTTATATGTTCCAGAAATCCGGTCTATATTA 2580
Qy 2101 ctccatgagacctcaaaaagctgtagaatgaatgaatgaatgaatgaatgaatgaat 2160
Db 2581 CTCCTATGACCTTCAAAAGCTAGATTAAGCAATGATGAGAGGCTTTTACACGAA 2640
Qy 2161 ttggaacatgctgtgagatataatgctgagatataatgaagaacacatctgatta 2220
Db 2641 TTTGACATGCTGTGATGATTTATGCTGATATCTATTTAGATTAAGCAATCTGATTA 2700
Qy 2221 gttacaatctcaaaaatctgataatcttaaggaagaaggagatataatctg 2280
Db 2701 GTTACAAATTCATAAAATTCATGATATTTTAAGGAAGGAGATTAATTAATTAAT 2760
Qy 2281 tatggagagacaatgaagcgaatcttctgagaagccttaagcttaatgcatctag 2340
Db 2761 TATGGAGAGACAATGAACGGGAATTTTTCAGAAAGCTTTAGTATGATGATCTAGC 2820
Qy 2341 gacacatgctgaagctttaaagctcaaaaatgctcgaaaaatctccaattataac 2400
Db 2821 GACCATGCTGAAGCTTTAAAGCTTCAAAAAATGCTCCGAAAACTTCCAAATTTATTA 2880
Qy 2401 gatcagatgaatctataatataatcaatca 2430
Db 2881 GATCAGATTAAGTCAATTAATTAATCAATTA 2910

RESULT 4
US-08-021-601-5
; Sequence 5, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Nichols, Peter J.
; APPLICANT: Atora, Naveen
; APPLICANT: Singh, Yogendra
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,601
; FILING DATE: 19930212
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414,057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
```

```
;
; LENGTH: 1368 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1368
; US-08-021-601-5

Query Match 31.4% Score 762.4 DB 1: Length 1368;
Best Local Similarity 99.2% Pred. No. 2.2e-146;
Matches 766: Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 100 gcgggcggtcagtgatgtagtgcgcgtaaaagagaagaagaagaagaagaagaaga 159
Db 1 GCGGGCGGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
Qy 160 aataagagaagaatgaagaacgaataaaacacagaagaagcatctaaagaaatcag 219
Db 61 AATAAGAGAAAGATGAAGAAAGAAATAAACACAGAAAGACATTTAAAGAAATCATG 120
Qy 220 aacacatgtaaaatgaagaatgaagaaggaggaagcgtttaaaagaaggcagcagaa 279
Db 121 AAACACATTTGTAATAATGAAGTAAAGGGAGAAAGCTTTAAAGAAAGAGCAGCAGAA 180
Qy 280 aagctactgagaagaatccatctgagctttagagatgataaaagcaatgagagaag 339
Db 181 AAGCTACTTGAAAGATGACCATCTGATGTTTAAAGATGATTAAGCAATGAGAGAAAG 240
Qy 340 atatatcttgatgagatgatatatacaaaacatatcttgaagaacatctcgaagat 399
Db 241 ATATATATTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Qy 400 aagaaaaaataaagaacattatggaagaatgcttataatcaatgaacattatglatat 459
Db 301 AAGAAAAAATTAAGACATTTATGGAAGATGCTTTTATTCATGACATTTATTTATAT 360
Qy 460 gcaaaagaagatatgaacccgtactgttaatccaaatctcgaagaatctatgaaat 519
Db 361 GCAAAAGAAAGATGACACCCGTACTTGTATCCAAATCTCGAAGATTTATGAAAT 420
Qy 520 actgaaaggacactgaagcttataatgaagaatgaagaatctatcaaggatcttca 579
Db 421 ACTGAAAGGACCTGACGTTTATTAAGAAATGATGATGATGATGATGATGATGATGAT 480
Qy 580 agtaaatatcaacatatagaatctttagatgataatcaatcaatcaatcaatcaat 639
Db 481 AGTAAATTAATCAACCATATCAGAAATTTTATGATGATGATGATGATGATGATGATGAT 540
Qy 640 tctgattcagatgacaagatcttataatcaatcaatcaatcaatcaatcaatcaat 699
Db 541 TCTGATTCAGATGACAGATCTTTTATTTACTAATCAAGCTTAAGCAATCCACAGAC 600
Qy 700 tttctctgagaatctcttgagaacaaatagcaatgaggtacaaagaatcttgagaag 759
Db 601 TTTTCTGTGAGATTTCTTGAACAAATATGCAATGAGTACAAAGATTTTTCGAAACT 660
Qy 760 ttgcatctataatgagacacagatcgtgagtttcaagccttataagcaggaaagct 819
Db 661 TTTGCAATTTATATGACGACACAGATCTGATGTTTATACCTTTTATGACCGAAGCT 720
Qy 820 tttaattcagatgaataatctaaagaagaagaataatctactcttgagaag 871
Db 721 TTTAATTCATGATGAATTTTAAACGACAAAGAAATTAATCTACGCGGAGC 772

RESULT 5
US-08-082-849B-5
; Sequence 5, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
```

APPLICANT: Leppia, Stephen H.
APPLICANT: Kilmpel, Kurt R.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
TITLE OF INVENTION: Related Methods
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1368 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1368
OTHER INFORMATION: /product= "LF(1-254)--TR--PE(401-602)"
US-08-082-849B-5
Query Match 31.4%; Score 762.4; DB 1; Length 1368;
Best Local Similarity 99.2%; Pred. No. 2.2e-146;
Matches 766; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

|||||
Db 241 ATATATTTTGTGATGTGATATTACAAAACATATATCTTTAGACGATTATCTGAAAGT 300
Qy 400 aagaaaaaaataaagacattatgggaagatgcttattacatgacattatglat 459
Db 301 AAGAAAAAATTAAGACATTTATGGGAAGATGCTTTATTACATGAACATTTATGTAT 360
Qy 460 gcaaaagaagatgatgacccgtaactgttaataccatcttcggaagattatgagaat 519
Db 361 GCAAAAAGAGATATGAACCCGTACTTTATATCCAACTTCGMAAGTTATGTAGAAAAT 420
Qy 520 accgaaaggcactgaagcttattatgaatagatgaatattatcaaggatattta 579
Db 421 ACTGAAAGGCACTGAACGTTTATATGAATAGTAAATTTATGAAGGATTTTAT 480
Qy 580 agtaaaataatcaaccatcagaaatttttagatgattaaataccatlaaaatgca 639
Db 481 AGTAAATTTATCAACCATTTAGAAATTTTATGATGTATTAATACCAATTAATAATGCA 540
Qy 640 tctgattcagatgacaaagcttttattactaatcagcttaaggacatccacagac 699
Db 541 TCTGATTCAGATGACAGATCTTTTATTTACTATATCAGCTTAAGAACATCCACAGAC 600
Qy 700 ttctcgtagaatctcttgacaacaaatagcaatgagttacaagaagatttcgaaagct 759
Db 601 TTTTCTGTAGAAATTTCTTGAAACAAATAGCATGAGTTACAGAAATTTGCAAAAGCT 660
Qy 760 ttgcataatatalcagacacagacatcgatgatttttaagctttatgacaggaaagct 819
Db 661 TTTGATATATATATGAGCAGACATGATGATTTTACAGCTTTATGACACCGGAAGCT 720
Qy 820 ttaattactatgataatttaacgaacaaagaataatattatcccttgagag 871
Db 721 TTTAATTATCATGATTAATTTAAGAACAAATAATATCTACTCGGCGAGC 772
RESULT 6
PCT-US94-01624-5
? Sequence 5, Application PC/TUS9401624
? GENERAL INFORMATION:
? APPLICANT: Leppia, Stephen H.
? APPLICANT: Kilmpel, Kurt R.
? APPLICANT: Arora, Naveen
? APPLICANT: Singh, Yogendra
? APPLICANT: Nichols, Peter J.
? TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
? NUMBER OF SEQUENCES: 31
? CORRESPONDENCE ADDRESSES:
? ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
? STREET: Stewart Street Tower, 20th Floor, One Market
? STREET: Plaza
? CITY: San Francisco
? STATE: CA
? COUNTRY: USA
? ZIP: 94105
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US94/01624
? FILING DATE: June 25, 1993
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: Weber, Kenneth A.
? REGISTRATION NUMBER: 31,677
? REFERENCE/DOCKET NUMBER: 15280-115
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415) 543-9600
? TELEFAX: (415) 543-5043
? INFORMATION FOR SEQ ID NO: 5:

Db 250 ATATATATTCTGATGATGATATTACAAAACATATATCTTTAGACGATTATCTGAGAT 309
Qy 400 aagaaaaaaataaagacattataggaaagatgctttatcatgaacttatgtat 459
Db 310 AAGAAAAAATAAAGACATTATGGAAAGAGCTTTATTAACATGAACATTTATGTATAT 369
Qy 460 gcaaaagaagatataaccgctactgttaaccatcttcggaagatatagtagaat 519
Db 370 GCAAAAAGAGATATGACCCGCTACTTGTATCCATCTTCGAGAGATTATGTAGAAAT 429
Qy 520 actgaaaggcactgacgtttctatgaaataggaatataatcaaggatattta 579
Db 430 ACTGAAAGGCACTGAACTTTATGTATGAATAGTATGATATTATCAAGGATATTTTA 489
Qy 580 agtaaaataataacacatacagaatctttagatgatatataacataaataatgca 639
Db 490 AGTAAATTTAATCAACCATATCAGAAATTTTGTATGATATTAAATCCATTAAAGATCA 549
Qy 640 tctgattcagatgacaagaatcttcttactaatcagcttaaggaaacatccacagac 699
Db 550 TCTGATTCAAGATGACAGATCTTTTATTACTATCAGCTTAAGAACATCCACAGAC 609
Qy 700 ttctctgtaaatcttctgaaacaaatagcaatgaggtacaaagaatatttgcgaagct 759
Db 610 TTTTCTGTAGAAATCTTGGAACAAATAGCAATGAGTACAGAAATTTGGGAAAGCT 669
Qy 760 ttgtcatattatatacgagccagacatcgtgattttaaagcttatgacccggaagct 819
Db 670 TTTGCAATTTATATCGAGCCACAGCATCGATGTTTACAGCTTATGACCGGAAAGCT 729
Qy 820 tttaatcacatgataaattttaaagcaagaagaataatctatc 863
Db 730 TTTAATTACATGATATAATTAAAGAAACAGAAATTAATCTAAC 773

RESULT 8
US-08-082-849B-7
; Sequence 7, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Kilmpel, Kurt R.
; APPLICANT: Aroza, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
; TITLE OF INVENTION: Related Methods
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/082,849B
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/021,601
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-161-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1425 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus anthracis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1416
; OTHER INFORMATION: /product= "Lf(1-254)--TR--PE(398-613)"
; US-08-082-849B-7

Query Match 31.4%; Score 762.4; DB 1; Length 1425;
Best Local Similarity 99.9%; Pred. No. 2; 2e-146;
Matches 763; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 100 gcgagcgatcgtgagtgatgacgttaaaagagaagaagaagaagaagaagaagaaga 159
Db 10 GCGGCGGCTCATGCTGATGTAGTATGCTACGTAAAGAGAAAGAGAAATTAAGATGAG 69
Qy 160 aataagaagaagaatgaagaacgaataaacaacagaagaagaatcttaagaagaatc 219
Db 70 AATAAGAGAAAGATGAGAAACGAAATTAACACAGGAGAGCATTTTAAAGGAATCATG 129
Qy 220 aaacacattttaaataagaaagaaagggggaagaagctttaaagaaggcgagaa 279
Db 130 AAACACATTTGTAATAATAGAAAGTAAGAGGAGAGAGCTGTTAAAGAGGCGACGAA 189
Qy 280 aagctacttgagaagaatgacatctgatttcttagaga tgataaagaagaatgagaaga 339
Db 190 AAGCTACTTGAGAAAGTACCATCTGATGTTTACAGATGTATAAGCAATTTGAGGAAG 249
Qy 340 atatatatttgatgattgatatataaacaatatatctttagaagaatattctgaagat 399
Db 250 ATATATTTGTGTGATGATATTAACAAACATATATCTTTAGAACGATTAATCGAAGAT 309
Qy 400 aagaaaaaaataaagaacattatagggaagaatgcttattacatgacaattatgatat 459
Db 310 AAGAAAAAATAAAGACATTTATGGAAAGAGCTTTATTACTGAACTTATGTATAT 369
Qy 460 gcaaaagaagatataaccgctactgttaaccatcttcggaagatatagtagaat 519
Db 370 GCAAAAAGAGATATGAACCCGCTACTTGTATCCATCTTCGAGAGATTATGTAGAAAT 429
Qy 520 actgaaaggcactgacgtttctatgaaataggaatataatcaaggatattta 579
Db 430 ACTGAAAGGCACTGAACTTTATGTATGAATAGTATGATATTATCAAGGATATTTTA 489
Qy 580 agtaaaataataacacatacagaatctttagatgatatataacataaataatgca 639
Db 490 AGTAAATTTAATCAACCATATCAGAAATTTTGTATGATATTAAATCCATTAAAGATCA 549
Qy 640 tctgattcagatgacaagaatcttcttactaatcagcttaaggaaacatccacagac 699
Db 550 TCTGATTCAAGATGACAGATCTTTTATTACTATCAGCTTAAGAACATCCACAGAC 609
Qy 700 ttctctgtaaatcttctgaaacaaatagcaatgaggtacaaagaatatttgcgaagct 759
Db 610 TTTTCTGTAGAAATCTTGGAACAAATAGCAATGAGTACAGAAATTTGGGAAAGCT 669
Qy 760 ttgtcatattatatacgagccagacatcgtgattttaaagcttatgacccggaagct 819
Db 670 TTTGCAATTTATATCGAGCCACAGCATCGATGTTTACAGCTTATGACCGGAAAGCT 729
Qy 820 tttaatcacatgataaattttaaagcaagaagaagaataatctatc 863
Db 730 TTTAATTACATGATATAATTAAAGAAACAGAAATTAATCTAAC 773

TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1524 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1524
US-08-021-601-9

Query Match 31.4%; Score 762.4; DB 1; Length 1524;
Best Local Similarity 99.9%; Pred. No. 2.2e-146;
Matches 763; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 100 gcggcggtcattgtagttagtgcacgttaaaagagaagaataaagtga 159
DB 1 GCGGGCGGTCATGATGTAGTATGACGTAAAGAGAGAAATTAAGATGAG 60
OY 160 aataagagaaaagatgaaacgaataaaacacaggaagcatttaaggaatcaty 219
DB 61 AATAAGAGAAAAGATGAGAAACGAATTAACACAGGAAGACATTTAAGGAATCATG 120
OY 220 aaacacattgttaaaatagaagtaaaaggaggagctgttaaaaaagagcagcagaa 279
DB 121 AAACACATTGTAAATAGAACTAAAGGGAGAGAACTGTTAAAAAGACGACAGAA 180
OY 280 aagctacttgagaagatccatctgatttttaagagatgataaagcaattggagaaag 339
DB 161 AAGCTACTTGAGAAAAGTACCATCTGATGTTTGAAGATGATTAAGCAATTGAGAGAAAG 240
OY 340 atatatctgtgacgtgacatactacaacacatactcttaagaagcattactcgaagat 399
DB 241 ATATATATTGTGATGATGATATTAACAAACATATATCTTTGAAGCATATCTGAAGAT 300
OY 400 aagaaaaataaaagaacatttaagggaagatgcttcttaacatgaaacattatgtatat 459
DB 301 AAGAAAAAATTAAGACATTTATGGAAGATGCTTTATATACATGACATTTATGATAT 360
OY 460 gcaaaagagaatgaaacccgtacttgaatcaactcttcgaagaatattagaanaat 519
DB 361 GCAAAAGAAAGATATGAACCCGCTACTTGTATTCATCTTCGAGAGATTATGAGAAAT 420
OY 520 actgaaaagcactgaacgtttatataagaaataggtaagatacttaacaggaatctta 579
DB 421 ACTGAAAAGGCCTGACGTTTATATGAATATAGTAAGATATTTATCAAGGATATTTTA 480
OY 580 agtaaaatcaatcaacacatacagaatctttagatgtaataaaccaataaaatgca 639
DB 481 AGTAAATTAATCAACCATATCAGAAATTTTATGATGATTAATTAATCAATTAATGCA 540
OY 640 tctgattgaatgagaagaatctttattactaactagcgttaagaagaatcccaagac 699
DB 541 TCTGATTAAGATGAGACAAAGATCTTTTATTTACTAATCACTTAAGAACATCCACAGAC 600
OY 700 ttctctgtaagaatctctgaaacaaatagcaatgaggtlaacgaagatttcgaaagct 759
DB 601 TTTTCTGTAGAAATCTTGTGSAACAAATATAGCATGAGTACAGAAATATTTCGAAAGCT 660
OY 760 ttgcatattatatacgaagccacagacatcgtagtcttcaagcttattgacccggaagct 819
DB 661 TTTGCAATTAATATCAGACACAGCATCGTATGTTTACACCTTTATACACCGAAGCT 720
OY 820 ttaataatcatgataaatttaacgaacacagaataaactctc 863
DB 721 TTTAATTACATGGATTAATTTTAACGACAGAAATTAATCTTAAC 764

RESULT 11
US-08-082-849B-9
Sequence 9; Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arore, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF INVENTION: Related Methods
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1524 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1524
OTHER INFORMATION: /Product= "LF(1-254) --TR--PE(362-613)"
US-08-082-849B-9
Query Match 31.4%; Score 762.4; DB 1; Length 1524;
Best Local Similarity 99.9%; Pred. No. 2.2e-146;
Matches 763; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 100 gcggcggtcattgtagttagtgcacgttaaaagagaagaataaagtga 159
DB 1 GCGGGCGGTCATGATGTAGTATGACGTAAAGAGAGAAATTAAGATGAG 60
OY 160 aataagagaaaagatgaaacgaataaaacacaggaagcatttaaggaatcaty 219
DB 61 AATAAGAGAAAAGATGAGAAACGAATTAACACAGGAAGACATTTAAGGAATCATG 120
OY 220 aaacacattgttaaaatagaagtaaaaggaggagctgttaaaaaagagcagcagaa 279
DB 121 AAACACATTGTAAATAGAACTAAAGGGAGAGAACTGTTAAAAAGACGACAGAA 180
OY 280 aagctacttgagaagatccatctgatttttaagagatgataaagcaattggagaaag 339

|||||
Db 181 AAGCTACTTGAGAAAGTACCTGATGTTTAAAGATGATTAAGCAATTGGAGGAAG 240
Qy 340 atacatactgtagtgcgatacacaacatactcttagaagcattccgaagc 399
Db 241 ATATATTTGGTGGATGGATTTACAAACATATATCTTTAGAACATTTCTGAAGAT 300
Qy 400 aagaaaaaaataaagacattttaggaaagatgcttattacatgaacattatglat 459
Db 301 AAGAAAAAATAAAAAGCATTTATGGAAGATGCTTTATTAACATTAAGATTAATATAT 360
Qy 460 gcaaaaagagatagaccgcttctgtaacatctcttggaagattatgtagaaat 519
Db 361 GCAAAAAGAGATATGAAACCGTACTGTAATCCAATCTTGGAAGATTAATGAAAT 420
Qy 520 actgaaagacgcgaacgcttattatgaagaaagtagaagatattacaaaggaatttta 579
Db 421 ACTGAAAAGCAGTCAACGTTTATTATGAATAGTAAGATATATCAAGGATATTTTA 480
Qy 580 agtaaaatlaatacaacatcagaatctttagatglatataacacattaaatgca 639
Db 481 AGTAAATTAATCAACCATATCAAGAAATTTTACATGATTAATTAACATTAATGA 540
Qy 640 tctgattcagatgtagaagatcttatttactaatacagcttaaggaacatccacagac 699
Db 541 TCTGATTCAGATGACACAGATCTTTTATTACTATCACTTAAGACATCCACAGAC 600
Qy 700 ttctcgtagaatcttggaaacaaatagcaatgaagtagaagaagatttgcgaagct 759
Db 601 TTTTCTGTAGAAATCTTGGAACAAATAGCAATGAGTACAAAGATTTGCGAAAGCT 660
Qy 760 ttgcatatcatatcagagccacagacatcgtagtctttagagcttattgacgcggaagct 819
Db 661 TTTGCAATTAATATACGAGCCACAGCATCGTATGCTTTTACAGCTTTATGACCGGAAGCT 720
Qy 820 tttaattacatgataaatttaacgaacagaataaataatcctatc 863
Db 721 TTTAATTACATGATTAATTAACGAACAGAAATTAATTAATCTAAC 764

RESULT 12
PCT-US94-01624-9
; Sequence 9, Application PC/TUS9401624
; GENERAL INFORMATION:
; APPLICANT: Leppia, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Airoa, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
; STREET: Steuart Street Tower, 20th Floor, One Market
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01624
; FILING DATE: June 25, 1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-115

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1524 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus anthracis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1524
; OTHER INFORMATION: /product=
; OTHER INFORMATION: "Lf(1-254)--Tr-PE(362-613)"
PCT-US94-01624-9

Query Match 31.4%; Score 762.4; DB 5; Length 1524;
Best Local Similarity 99.9%; Pred. No. 2,3e-146;
Matches 763; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 100 gggggcgatcagtgatgtagtgcagtaaaagagaagaaataaagatgag 159
Db 1 GCGGGCGGTCAATGCTGATGTGATGATGACCTAAAGAGAAAGAAATTAAGATGAG 60
Qy 160 aataagaaaagatgtagaagcaacaaataaacaaggaagagcattaaagaaatcag 219
Db 61 AATTAAGAAATTAATGAGAAAGCAATTAACACAGAAAGCATTTTAAGAAATCAATG 120
Qy 220 aaacacattgtaaaatagaagtaaaagggaggaagctgttaaaaaagagcagcagaa 279
Db 121 AAACACATTTGTAATAATGAAGTAAAGGGGAGGAGAGCTGTTAAAAAGAGCAGCAAGAA 180
Qy 280 aagctaacttgagaagatcacatcgtatgctttagagatgtaaaagcaattggaagaa 339
Db 181 AAGCTACTTGAGAAAGTACCATCTGATGCTTTTACAGATGATTAAGCAATTCGAGGAAG 240
Qy 340 atacatactgtagtgcgatacacaacatactcttagaagcattcctgaagc 399
Db 241 ATATATTTGGTGGATGGATTTACAAACATATATCTTTAGAACATTTATCTGGAAGAT 300
Qy 400 aagaaaaaaataaagacattttaggaaagatgcttattacatgaacattatglat 459
Db 301 AAGAAAAAATAAAAAGCATTTATGGAAGATGCTTTATTAACATTAATTAATGATTAAT 360
Qy 460 gcaaaaagagatagaccgcttctgtaacatctcttggaagattatgtagaaat 519
Db 361 GCAAAAAGAGATATGAAACCGTACTGTAATCCAATCTTGGAAGATTAATGGAAT 420
Qy 520 actgaaagacgcgaacgcttattatgaagaaagtagaagatattacaaaggaatttta 579
Db 421 ACTGAAAAGCAGTCAACGTTTATTATGAATAGTAAGATATATCAAGGATATTTTA 480
Qy 580 agtaaaatlaatacaacatcagaatctttagatglatataacacattaaatgca 639
Db 481 AGTAAATTAATCAACCATATCAAGAAATTTTACATGATTAATTAACATTAATGA 540
Qy 640 tctgattcagatgtagaagatcttatttactaatacagcttaaggaacatccacagac 699
Db 541 TCTGATTCAGATGACACAGATCTTTTATTACTATCACTTAAGACATCCACAGAC 600
Qy 700 ttctcgtagaatcttggaaacaaatagcaatgaagtagaagaagatttgcgaagct 759
Db 601 TTTTCTGTAGAAATCTTGGAACAAATAGCAATAGTACAAAGATTTTGGCAAAAGCT 660
Qy 760 ttgcatatcatatcagagccacagacatcgtagtctttagagcttattgacgcggaagct 819
Db 661 TTTGCAATTAATATACGAGCCACAGCATCGTATGCTTTTACAGCTTTATGACCGGAAGCT 720

Qy 820 tttaatacatgataaatttaacgacagaataaataatc 863
Db 721 TTTAATTACATGATTAATTTAACGACAGAAATAATTAATCTAAC 764

RESULT 13
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/232.463
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300. 6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29, 768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F15
US-08-232-463-14

Query Match 3.8%; Score 91.6; DB 1; Length 7218;
Best Local Similarity 2.6%; Pred. No. 3.9e-10;
Matches 10; Conservative 255; Mismatches 119; Indels 0; Gaps 0;
Qy 89 ttgtacaggggcgggcgtcattgtgtatgtatgcacgttaaaagaagaaga 148
Db 1440 TCGTACRR 1381
Qy 149 ataagaatgagataagaagaagaatgaagaacgaataaacaacaggaagcatt 208
Db 1380 RRR 1321
Qy 209 aggaatcatgaacacattgttaaaataagaagtaaaaggaggagagctgttaaa 268
Db 1320 RRR 1261
Qy 269 aggcagcagaaagctactgtgaaagtaacacatcgtatgtttaagagatlaaagc 328
Db 4318

Db 1260 RRR 1201
Qy 329 ttgaggaagaatataatgtgtgagtgatctcaaacataatcttgaagcat 368
Db 1200 RRR 1141
Qy 389 tatctgaagaataagaataaagacattatggaagaatgtcttataatcagac 448
Db 1140 RRR 1081
Qy 449 atatgtatatacgaagaagaat 472
Db 1080 RRR 1057

RESULT 14
US-08-973-462-2
; Sequence 2, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUHLHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973.462B
; EARLIER FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5361
; TYPE: DNA
; ORGANISM: P. falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(5361)
US-08-973-462-2

Query Match 2.7%; Score 66.4; DB 4; Length 5361;
Best Local Similarity 42.6%; Pred. No. 4.7e-05;
Matches 468; Conservative 0; Mismatches 621; Indels 9; Gaps 2;
Qy 110 atgtgtatgttaggtatgcacgttaaaagaagaagaataaagaatgaataaga 169
Db 3902 atgaagtgttagatataaagaatgtcgaagaagaacgaatcgaaagaatctgtatt 3961
Qy 170 aagaatgaagaacgaataaacaacaggaagagcatttaaaagaaatcatgaacac 229
Db 3962 aagatcttgaagaagatatattaaagaagtaaaagaacaaacgaactgtgaag 4021
Qy 230 taaaataagaatgaagaaggagagcgtgttaaaagaaggagcaagaagaactctg 289
Db 4022 ttcttaagaatataaagaatataaacaactatgaacgaatatttaagaaga 4081
Qy 290 agaaaatcacatctgttctttagaagtataaagaacattgtgaggaagatata 349
Db 4082 aatagaagaagaatcttctgaaataatcgaaagaagaagcgtgaagaataaagctc 4140
Qy 350 ttgattgtatatacaaaaacataatctttagaagcattatctgaagaataa 409
Db 4141 -gaagcaatataataaagaagtaatcttcatctagaagttgaagaaga 4198
Qy 410 taaagaacattttgggaagaatgtcttattacatgaacattatgtatgtcgaaga 469
Db 4199 aagaagtaacagaaataaagaagaagtgaaacataataaagtggtgagcgcatata 4258
Qy 470 gatataaccgctactgttaaatcaatcttcgaaagaatattatgagaataacgaaag 529
Db 4259 aaggttggaaagaagatgtattgaagaagtgatgatttaaaagaaagtattatgaca 4318

Oy	550	cactgaaagcttattatcatgaataaggttaagaatatcttacaaaggagatacttttaagttaaatta	589
Db	4319	tgttaaaaggagatattgtaatttttaggttatgatagaataaggaagcttttagaagatgttaacaa	4378
Oy	590	atcaaccatcagaataatttttagtgttgtaataaccaattaaatgcattgttcag	649
Db	4379	caaaaacctggaaaaagcttgaatcccttaaaagatgctttactctgttcattgaagcatbg	4438
Oy	650	atggacaagatcttttacttacttaacgccttaaggaacatcccacagacttctcgtag	709
Db	4439	atgaaagaacaaatgaaaacaaagaaaaaacgcacaagacttaagttgtaagaagattat	4498
Oy	710	aattcttgaaacaaattatgcaatggcgatcgacaaagattctcgaaagcttttgcatatt	769
Db	4499	taaaagaagaaggtttaaagaagaacaaagaaaaatacaaaaaaagaagaagtttg	4558
Oy	770	ataticgaagccaagatcgtgtgatgttttaacgctttatlgcacgggaagcttttaataca	829
Db	4559	atatttaagatattgaaccacaaagattgaatatgtaagttgaaatgaaatgaaatgtaagata	4618
Oy	830	tgtgataaatcttaacgaacaaagaaataaatctctccttggaaagaacttaagaatcaacgga	889
Db	4619	tagaagaagatgtgaagaagaatatalagaagaagataatagaagaataaagtttgaagata	4678
Oy	890	tgcctgcagaatattgaanaaatgysaaagaataaacaagacatcatacacctgagcgatt	949
Db	4679	t-----agatgaagatataagttgaagatatatgtggaagacaaagaatgaagttatagatt	4732
Oy	950	ctttactgtgaagaagaagaggaactttaaaaaagctgcagattccctattgsgccaaga	1009
Db	4733	taatgtccaaaagaagaaacgcactlgtaaaagttlaaagcgaagaaaaaataattagaaa	4792
Oy	1010	aagatgacataattcatcttcttattctccaagaagaagaagagcttctaanaagaatatacaa	1069
Db	4793	aaaaagcttgaagaagagtgcttagtgctttaaaaaacacgtagacgaagtaagaataatag	4852
Oy	1070	ttgatagtatgtatctttatctactctgagaaaagagtttttaaaaaagctacaatatg	1129
Db	4853	ttcaaaaaatgataagaagattgataagaagtatctaagctttagatcaaaaaatg	4912
Oy	1130	atactcgatcttcttattctgtaagaagaagaagaagctttaaataatgatacaggtgata	1189
Db	4913	atgttaaccaatgttttaaaaaacaaatacaagattcttttagtaagttaaaaacttcgttaa	4972
Oy	1190	gtagtaatccttactcgt	1207
Db	4973	aaaaataaagatttg	4990
RESULT 15			
US-08-973-462-1			
Sequence 1, Application US/08973462B			
Patent No. 6191270			
GENERAL INFORMATION:			
APPLICANT: DROUHE, PIERRE			
APPLICANT: DAUBERSTES, PIERRE			
TITLE OF INVENTION: MLARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES			
FILE REFERENCE: 0660-0125-0 PCT			
CURRENT APPLICATION NUMBER: US/08/973, 462B			
CURRENT FILING DATE: 1998-02-06			
EARLIER APPLICATION NUMBER: PCT/FR96/00894			
EARLIER FILING DATE: 1996-06-12			
EARLIER APPLICATION NUMBER: FR 95/07007			
EARLIER FILING DATE: 1995-06-13			
NUMBER OF SEQ ID NOS: 29			
SOFTWARE: PatentIn Ver. 2.0			
SEQ ID NO 1			
LENGTH: 6152			
TYPE: DNA			
ORGANISM: P. falciparum			
US-08-973-462-1			

Query Match	2.7%	Score 65.4	DB 4	Length 6152
Best Local Similarity	42.6%	Pred. No. 4.8e-05		
Matches 468	Conservative 0	Mismatches 621	Indels 9	Gaps 2
QY	110	atgtgcatgttaagtatgcgcgttaaaagagagaaagataaagtatgaataagaa	169	
DB	4146	atgaagtctgttaaatataaaagatgttcgaagagacaaagatcgaaaaagatctgtttaa	4205	
QY	170	aagatgaagaagaaataaacaacgcggaagagcatttaaggaatactatgaacacatg	229	
DB	4206	aagatcttgaagaagatcatatataaagaagataaagaataccaagaacttgaagtga	4265	
QY	230	taaaatagaagaataaaggaggaggagcgtgttaaaaaagagcgacgaaaaagcctactg	289	
DB	4266	ttcttgaagaagtataaagaatataaaacatttgaacacgataatttctgaaggaataaag	4325	
QY	290	agaagatccatctgatgttctttagagatgtataaaagcaatttggaggaagatataatg	349	
DB	4326	aaatagaaaaagatcatctttagaaaaattccgaagaagaagcctgaagaataaaagctt-	4384	
QY	350	tgtatgtgtatatactcaaaacataatcttgaagacattatctgaagaataaagaa	409	
DB	4385	--gaagcagatatataaaagaagatactctctcttgaagcttgaagaagaaaaaaattag	4442	
QY	410	taaaagacattatggaggaagaatgtcttaatacgaagaacattatcatatgaagaaag	469	
DB	4443	aagaaagtacagaaatataaaagaagggtcgaacacataataaagtgtgtatgctgcatata	4502	
QY	470	gatatgaaccgctactctgttaatcacaacttcggaagatattgtagaataactgtgaag	529	
DB	4503	aaggtcttgaagaagaatgattttagaagaagtagatgactttaaagaagtattataga	4562	
QY	530	cactgaagcttattatgaataagtgatgatatcttaaggatatttttaagtaata	589	
DB	4563	tgttaagggagatctgtgaatcagggtatagataaggaagaagctttagaagaatgaaca	4622	
QY	590	atcacacatatacgaaatctttagatgtataatacacaatlaaaactcatctgtatcag	649	
DB	4623	caaaacttggagaagaaggttgaactccttaaaagatgttctatctatgctatgaagcattg	4682	
QY	650	atggacgaagatctttatttactaatcagcttaaggaaacatcccaagactttctgtag	709	
DB	4683	atgaaagaacaaatgaaaaacaagaaaaaagctccaagaagcctaagtttgaagaagatatat	4742	
QY	710	aattcttgaacaacaatatgcaatgtgaatgaagaagatatttgcgaagctttgtcatat	769	
DB	4743	taaaagaagaaggtttaaagaagaacaaagaaaaaataacaaaaaagaagaagtgttg	4802	
QY	770	atatacgaagccacagacatcgttgaatgttatacagcttattatgcacgcggaagctttaata	829	
DB	4803	atattaaagatatagaaccaaagaatgaatatagtagaagcttgaatgtgaagaatgtgaata	4862	
QY	830	tgtataaatttaacgaacaagaataaatacttactcttgaagaataacttaagaatcaagga	889	
DB	4863	tagaagaagaatgttaagaataagaatacagaagaatataagaagaatacaaaagtctgaagata	4922	
QY	890	tgtcttcaagatatcaaaaatcggaaagaagataaaacgcagcatatacaacacttgaagcatt	949	
DB	4923	t-----agatgaagatatagtagtgaatgatctgtgtgaagacaaagatgaagtatatagatt	4976	
QY	950	ctttatctgaagaaggaagaggaacttttaaaaagctgcagatcttcattatgagccaaga	1009	
DB	4977	tatatgtccaaaanaagaaacgcatttgaaaaagtttaagcggaaaaaagaaaaattatgaaa	5036	
QY	1010	aagatgaataatcatcttcttatactcgaagaagaagaagagcttcttaaaagaatatcaaa	1069	
DB	5037	aaaaggttgaagaagaagtgtagtgtgtcttaaaaaacogtagaagaaagttaagaatactg	5096	
QY	1070	ttgatagtatgtatttttattatctactctgggaaaaagattttaaanaagctatacaatgt	1129	
DB	5097	ttcaaaaatttataaagaagtttgaataaagaatctcaaaacttataaatcaaaaata	5156	

Mon' Dec 3 10:23:17 2001

Qy 1130 atcttcgtagcttcttcttcgaagaagaaagagcttttaattcgaatacaggtagata 1189
 Db 5157 atgcttctaattgcttttaacacaaacaaacgaatttttttagtaagtaataaactctgtaa 5216
 Qy 1190 gttagtaactcttattctcg 1207
 Db 5217 aaaaataaaaagctattctg 5234

Search completed: December 2, 2001, 13:51:02
Job time: 117 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 2, 2001, 13:49:05 ; Search time 129.57 Seconds
(without alignments)
16078.570 Million cell updates/sec

Title: US-09-747-521-1
Perfect score: 2430
Sequence: 1 atgataataaaaaagaatt.....agttcattatcaactacataa 2430

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N.Geneseq_1101:*
2: /SID2/gcgdata/geneseq/geneseq/NA1980.DAT:*
3: /SID2/gcgdata/geneseq/geneseq/NA1981.DAT:*
4: /SID2/gcgdata/geneseq/geneseq/NA1982.DAT:*
5: /SID2/gcgdata/geneseq/geneseq/NA1983.DAT:*
6: /SID2/gcgdata/geneseq/geneseq/NA1984.DAT:*
7: /SID2/gcgdata/geneseq/geneseq/NA1985.DAT:*
8: /SID2/gcgdata/geneseq/geneseq/NA1986.DAT:*
9: /SID2/gcgdata/geneseq/geneseq/NA1987.DAT:*
10: /SID2/gcgdata/geneseq/geneseq/NA1988.DAT:*
11: /SID2/gcgdata/geneseq/geneseq/NA1989.DAT:*
12: /SID2/gcgdata/geneseq/geneseq/NA1990.DAT:*
13: /SID2/gcgdata/geneseq/geneseq/NA1991.DAT:*
14: /SID2/gcgdata/geneseq/geneseq/NA1992.DAT:*
15: /SID2/gcgdata/geneseq/geneseq/NA1993.DAT:*
16: /SID2/gcgdata/geneseq/geneseq/NA1994.DAT:*
17: /SID2/gcgdata/geneseq/geneseq/NA1995.DAT:*
18: /SID2/gcgdata/geneseq/geneseq/NA1996.DAT:*
19: /SID2/gcgdata/geneseq/geneseq/NA1997.DAT:*
20: /SID2/gcgdata/geneseq/geneseq/NA1998.DAT:*
21: /SID2/gcgdata/geneseq/geneseq/NA1999.DAT:*
22: /SID2/gcgdata/geneseq/geneseq/NA2000.DAT:*
23: /SID2/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2430	100.0	2430	22	AAC86015	Wild type B. anthr
2	2430	100.0	3284	15	AAO70179	Sequence encoding
3	767.6	31.6	1455	22	AA500250	Lfn-Bcl-Xl apolpos
4	762.4	31.4	1368	15	AAO70181	Lfn(1-254)-TR-PE
5	762.4	31.4	1425	15	AAO70182	Lfn(1-254)-TR-PE
6	762.4	31.4	1524	15	AAO70183	Lfn(1-254)-TR-PE
7	211.2	8.7	936	22	AA58252	Oligonucleotide D1
8	211.2	8.7	936	22	AA58254	Oligonucleotide D1
9	211.2	8.7	936	22	AA58257	Oligonucleotide D1
10	211.2	8.7	936	22	AA58259	Oligonucleotide D1
11	211.2	8.7	936	22	AA58262	Oligonucleotide D2

C	12	211.2	8.7	938	22	AA58255	Oligonucleotide D1
	13	209.4	8.6	936	22	AA58252	Oligonucleotide D1
	14	209.4	8.6	936	22	AA58254	Oligonucleotide D1
	15	209.4	8.6	936	22	AA58257	Oligonucleotide D1
	16	209.4	8.6	936	22	AA58259	Oligonucleotide D1
	17	209.4	8.6	936	22	AA58262	Oligonucleotide D2
	18	209.4	8.6	938	22	AA58255	Oligonucleotide D1
	19	157.8	6.5	2990	11	AAO04123	Adenyl cyclase gen
	20	98.4	4.0	1686	16	AAO87587	DNA encoding Leuco
	21	96	4.0	3399	17	AAO5868	Chicken leucocytos
	22	83.2	3.4	6644	20	AAO33181	Base sequence of t
	23	83.2	3.4	7372	20	AAO33182	Base sequence of t
	24	83.2	3.4	7797	20	AAO33180	Cowpox virus bsr f
	25	83.2	3.4	7996	20	AAO33183	Base sequence of t
	26	82.6	3.4	5940	21	AAO70105	Plasmodium falcipa
	27	82	3.4	1998	21	AAO70105	Plasmodium falcipa
	28	80.8	3.3	4590	7	AAO60472	Sequence encoding
	29	75.2	3.1	3579	21	AAO70099	Plasmodium falcipa
	30	74.6	3.1	3549	21	AAO70223	Plasmodium falcipa
	31	74.6	3.1	5340	22	AAO26289	P. falciparum eg5
	32	73.6	3.0	1527	21	AAO70121	Plasmodium falcipa
	33	72.4	3.0	3927	21	AAO70101	Plasmodium falcipa
	34	72.2	3.0	3567	21	AAO70117	Plasmodium falcipa
	35	71	2.9	1035	21	AAO1936	Human colon cancer
	36	71	2.9	3744	21	AAO70149	Plasmodium falcipa
	37	70.8	2.9	7326	21	AAO70095	Plasmodium falcipa
	38	70	2.9	7458	21	AAO70106	Plasmodium falcipa
	39	69.2	2.8	10640	22	AAO03729	P. falciparum telo
	40	69	2.8	4468	21	AAO62126	DNA encoding a cal
	41	69	2.8	4611	21	AAO62127	Sequence encoding
	42	68.4	2.8	1612	7	AAO0392	Oligonucleotide D1
	43	67.4	2.8	244	22	AAO58238	Plasmodium falcipa
	44	67	2.8	3837	21	AAO70211	Plasmodium falcipa
	45	66.8	2.7	5994	21	AAO70222	Plasmodium falcipa

ALIGNMENTS

RESULT	1
ID	AAC86015 standard; cDNA; 2430 BP.
AC	AAC86015;
DT	29-AUG-2001 (first entry)
DE	Wild type B. anthracis lethal factor coding sequence.
XX	
KW	Lethal factor; LF; immunogen; LF4; protective antigen; PA; DNA vaccine;
KM	humoral; cell-mediated; immune memory response; ss.
XX	
OS	Bacillus anthracis.
XX	
FH	Key
FT	sig-peptide
FT	1..99
FT	Location/Qualifiers
FT	/*tag= a
FT	mat-peptide
FT	100..2430
FT	/*tag= b
FT	/product= "LF"
FT	124..855
FT	/*tag= c
FT	/product= "Encodes LF4 peptide"
XX	
XX	WO200145639-A2.
PN	
XX	28-JUN-2001.
XX	
PD	
XX	
XX	21-DEC-2000; 2000WO-US4912.
PF	
XX	
XX	22-DEC-1999; 99US-0171459.
PR	
XX	
XX	
PA	(OHIS) UNIV OHIO STATE RES FOUND.

PA (GALL/) GALLOWAY D R.
PA (MATE/) MATECZUN A J.
XX Galloway DR, Mateczun AJ;
XX
XX
DR WPI: 2001-408540/43.
DR P-PSDB: AAB47305.
XX
XX
PT protecting animal against lethal infection with Bacillus anthracis, by
PT administering wild type or mutated form of Bacillus anthracis lethal
PT factor protein or its fragment or a nucleic acid encoding the mutated
PT protein -
PS
PS Disclosure: Fig 1: 33pp: English.
XX
XX This sequence encodes the B. anthracis lethal factor (LF). An
CC immunogenic fragment of LF, LF4, can be used to produce an immune
CC response which protects an animal against lethal infection with
CC Bacillus anthracis. DNA encoding the B. anthracis LF can be used
CC in conjunction with DNA encoding the protective antigen (PA) in a
CC DNA vaccine. Using a DNA vaccine which encodes the mutated LF protein
CC or fragment alone or in combination with a DNA encoding the PA protein
CC or its fragment, both components (humoral and cell-mediated) of the
CC immune system are stimulated, which results in longer term immune
CC memory response. The combined use of a mutated LF and PA gene or their
CC fragments results in a higher level of immune response, as judged by
CC overall serum antibody titers for LF and PA antigens, than the use of
CC either LF or PA genes in separate immunizations.
XX
XX Sequence 2430 BP: 985 A; 296 C; 445 G; 704 T; 0 other;

Query Match 100.0%; Score 2430; DB 22; Length 2430;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 2430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgaataataaaagaattataaagaatgaatgaatgctggtttagtaacagaatt 60
DB 1 atgaataataaaagaattataaagaatgaatgaatgctggtttagtaacagaatt 60
QY 61 acttgagtggtccgcttcttacccttgtaacagggcgggcggtcatgtgtatga 120
DB 61 acttgagtggtccgcttcttacccttgtaacagggcgggcggtcatgtgtatga 120
QY 121 ggtatgcaatgaagaagaagaagaagaagaagaagaagaagaagaagaaga 180
DB 121 ggtatgcaatgaagaagaagaagaagaagaagaagaagaagaagaagaaga 180
QY 181 cgaataaaacacaggaagaagatttaagaagaatcgtgaacacgtttaaataaga 240
DB 181 cgaataaaacacaggaagaagatttaagaagaatcgtgaacacgtttaaataaga 240
QY 241 gtaaaaggagggaagctgttaaaagaagcagcagaagaacttacttgaagaatga 300
DB 241 gtaaaaggagggaagctgttaaaagaagcagcagaagaacttacttgaagaatga 300
QY 301 tctgagtgttttagagatgtataaagaatctggagaagaatataatctgtgagtgt 360
DB 301 tctgagtgttttagagatgtataaagaatctggagaagaatataatctgtgagtgt 360
QY 361 attcaaaaacatatacttttagaagaatatacgtgaagaataaagaataaagaatc 420
DB 361 attcaaaaacatatacttttagaagaatatacgtgaagaataaagaataaagaatc 420
QY 421 tatgggaagaatgctttattacatgaacattatgtatcgtcaaaagaagaatgaacc 480
DB 421 tatgggaagaatgctttattacatgaacattatgtatcgtcaaaagaagaatgaacc 480
QY 481 gtaactgttaaccatcttcggaagaattatgtagaanaatacgtgaagaagcactga 540
DB 481 gtaactgttaaccatcttcggaagaattatgtagaanaatacgtgaagaagcactga 540
QY 541 tattatgaataatgtaagaatataatcaaggatattttaagtaaaataatcaacatat 600

DB 541 tattatgaataatgtaagaatataatcaaggatattttaagtaaaataatcaacatat 600
QY 601 cagaatctttagatgtataataatcattaaagaagcctgattcagatgagaagaat 660
DB 601 cagaatctttagatgtataataatcattaaagaagcctgattcagatgagaagaat 660
QY 661 ctttattactaatcagcttaaggaacatccacagacttctctgtaaatctgtga 720
DB 661 ctttattactaatcagcttaaggaacatccacagacttctctgtaaatctgtga 720
QY 721 caaaatagaagaatgagtaagaagaatatttggaagaacttttgcatatatcgagca 780
DB 721 caaaatagaagaatgagtaagaagaatatttggaagaacttttgcatatatcgagca 780
QY 781 cagcatcgatgattttacagctttatgacccgaagaacttttaattacatgaataatc 840
DB 781 cagcatcgatgattttacagctttatgacccgaagaacttttaattacatgaataatc 840
QY 841 aacgaacaagaataataatctatctctgtgaagaacttaagaatcaacgaatgctga 900
DB 841 aacgaacaagaataataatctatctctgtgaagaacttaagaatcaacgaatgctga 900
QY 901 tatgaanaatgggaagaagaataaagcactatcaacactgtgaagcattcttatcga 960
DB 901 tatgaanaatgggaagaagaataaagcactatcaacactgtgaagcattcttatcga 960
QY 961 gaagaagaagccttttaaaagaagcctgacatccatcgtgaacccaagaagaatgaca 1020
DB 961 gaagaagaagccttttaaaagaagcctgacatccatcgtgaacccaagaagaatgaca 1020
QY 1021 attcatctttatctcaagaagaagaagccttcaaaagaataacatatgataagtagt 1080
DB 1021 attcatctttatctcaagaagaagaagccttcaaaagaataacatatgataagtagt 1080
QY 1081 gatttttttctacgtgagaagaagaagcttttaaaagaactcaaatgatatctgtgt 1140
DB 1081 gatttttttctacgtgagaagaagaagcttttaaaagaactcaaatgatatctgtgt 1140
QY 1141 tctttatctgaagaagaagaagaagccttttaaaagaatcagaatcaggtatgaatcc 1200
DB 1141 tctttatctgaagaagaagaagaagccttttaaaagaatcagaatcaggtatgaatcc 1200
QY 1201 ttatctgaaaagaagaagaagcttttaaaagaagcgtgaactgtatatcaacatatgt 1260
DB 1201 ttatctgaaaagaagaagaagcttttaaaagaagcgtgaactgtatatcaacatatgt 1260
QY 1261 attaatcaaaagcttgcaagaatcagaagaaggttaattgtagtccgtcaaatcaatc 1320
DB 1261 attaatcaaaagcttgcaagaatcagaagaaggttaattgtagtccgtcaaatcaatc 1320
QY 1321 gtaagaagaagcagatataaagaagatctcaaaatattgtacttatacatcaatcatt 1380
DB 1321 gtaagaagaagcagatataaagaagatctcaaaatattgtacttatacatcaatcatt 1380
QY 1381 ggaagtaacctgtacataaanaatttatctgtatgaaataatgaatacaataaacttaca 1440
DB 1381 ggaagtaacctgtacataaanaatttatctgtatgaaataatgaatacaataaacttaca 1440
QY 1441 ggaacccctagtggtgagatttagtattccactgtgaataacaaataataatgaggtatt 1500
DB 1441 ggaacccctagtggtgagatttagtattccactgtgaataacaaataataatgaggtatt 1500
QY 1501 ttcaatgaatccaanaaanaattccaataatgatttccagtaacatataatgtgtgtat 1560
DB 1501 ttcaatgaatccaanaaanaattccaataatgatttccagtaacatataatgtgtgtat 1560
QY 1561 ataaatgaaaggcctgtcattagataatgagcgtttgaaatggaagaatccaatlatacca 1620
DB 1561 ataaatgaaaggcctgtcattagataatgagcgtttgaaatggaagaatccaatlatacca 1620
QY 1621 gatactcgagcagatatttaagaanaatggaagcctatattcaacaagaagaactcggtcg 1680

Db	1621	gatactcgcgcgagataatttagaanaatggaagcctatattctcaagaanaacatgcgtctg	1660
Qy	1661	gaataaaggatgycacaanaalttaagcgaatccggaanaagaatataaaggattgatcg	1740
Db	1661	gaataaaggatgycacaanaalttaagcgaatccggaanaagaatataaaggattgatcg	1740
Qy	1741	aaagtatgcccgaaggtaaataatagatatacaaaaattccaagaagccgttaataaat	1800
Db	1741	aaagtatgcccgaaggtaaataatagatatacaaaaattccaagaagccgttaataaat	1800
Qy	1801	caggaaatggaataaagatattaggtttaccgaataatcacaaagcttattacattcaactg	1860
Db	1801	caggaaatggaataaagatattaggtttaccgaataatcacaaagcttattacattcaactg	1860
Qy	1861	cataatagatatgcattcccaatattctgtagaagaatgcttattatataatggaatgtagaa	1920
Db	1861	cataatagatatgcattcccaatattctgtagaagaatgcttattatataatggaatgtagaa	1920
Qy	1921	aataataatcaaaagtgtactcttataaaaaaggtaacaaattacttctgtgtatgtagaa	1960
Db	1921	aataataatcaaaagtgtactcttataaaaaaggtaacaaattacttctgtgtatgtagaa	1960
Qy	1961	agattgtttttaccggtattactctcccaataatagatgcgaacaatatcacatcaaat	2040
Db	1961	agattgtttttaccggtattactctcccaataatagatgcgaacaatatcacatcaaat	2040
Qy	2041	gagatataatgagcaagtctcatcaaaaaggttatatgtttccgaatcccggttctatat	2100
Db	2041	gagatataatgagcaagtctcatcaaaaaggttatatgtttccgaatcccggttctatat	2100
Qy	2101	ctccatggaaccttcaaaaaggtgtagaataaaggaatgatatgtaggtttatacagaa	2160
Db	2101	ctccatggaaccttcaaaaaggtgtagaataaaggaatgatatgtaggtttatacagaa	2160
Qy	2161	tttgagaatgctgtggtgtagttatgctgtgatatctattagataaagaacaaatctgatta	2220
Db	2161	tttgagaatgctgtggtgtagttatgctgtgatatctattagataaagaacaaatctgatta	2220
Qy	2221	gttcaaatctctaaaaaatcatctgtatatttttaagaagaagggagtaatttaactcg	2280
Db	2221	gttcaaatctctaaaaaatcatctgtatatttttaagaagaagggagtaatttaactcg	2280
Qy	2281	tatgggaggaacaatgtagcggaaatcttttcagaagccttaggttaatgcatctacg	2340
Db	2281	tatgggaggaacaatgtagcggaaatcttttcagaagccttaggttaatgcatctacg	2340
Qy	2341	gaccatgcctgaacgttttaaaaagttcaaaaaaagtcgcgaagaacttccaattttat	2400
Db	2341	gaccatgcctgaacgttttaaaaagttcaaaaaaagtcgcgaagaacttccaattttat	2400
Qy	2401	gaccagattaaagttcatctataactcataa	2430
Db	2401	gaccagattaaagttcatctataactcataa	2430
RESULT 2			
AAO70179	ID	AAO70179 standard; DNA; 3294 BP.	
XX	AC	AAO70179;	
XX	DT	03-APR-1995 (first entry)	
DE		Sequence encoding lethal factor of Bacillus anthracis.	
XX		Anthrax; Bacillus anthracis; fusion protein; lethal factor;	
KW		protective antigen; cell killing; targeting; targeting; pathogen;	
KW		intracellular; HIV; human immunodeficiency virus; toxin; ss.	
XX		Bacillus anthracis.	
XX			
XX	Key	Location/Qualifiers	
FT	CDS	580..2910	

```

FT      /*tag= a
XX      /product= Anthrax lethal factor.
XX
XX      WO9418332-A.
XX
XX      18-AUG-1994.
XX
XX      PF      14-FEB-1994; 94WO-US01624.
XX
XX      PR      12-FEB-1993; 93US-0021601.
XX      PR      25-JUN-1993; 93US-0082849.
XX
XX      PA      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX      PI      Arora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;
XX      DR      WPI: 1994-279753/34.
XX      DR      P-PSDB; AAR60178.
XX
XX      Nucleic acid encoding anthrax toxin fusion protein - useful for
XX      PT      targeting toxin to specific cells, eg for killing tumour cells
XX      PT      or HIV-infected cells
XX
XX      PS      Disclosure: Page 72-75; 124pp; English.
XX
XX      The sequence encoding the lethal factor of Bacillus anthracis may be
XX      CC      used in the construction of a nucleic acid which encodes a fusion
XX      CC      protein comprising the anthrax protective antigen binding domain of
XX      CC      the native anthrax lethal factor and a sequence encoding an activity
XX      CC      inducing domain of a second protein. The fusion proteins are useful
XX      CC      for the specific killing of tumour cells or the killing of cells
XX      CC      infected with intracellular pathogens, especially HIV.
XX
XX      SO      Sequence 3294 BP; 1335 A; 391 C; 564 G; 1004 T; 0 other;

Query Match      100.0%; Score 2430; DB 15; Length 3294;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 atgaatatataaaagaattataaagaatagtatctgttttagtaacgaactt 60
DB      481 atgaatatataaaagaattataaagaatagtatctgttttagtaacgaactt 540
QY      61 actttgagtggtcccgctcttaccgcccttgtaacaggggcgggcggtcatgtgatgta 120
DB      541 actttgagtggtcccgctcttaccgcccttgtaacaggggcgggcggtcatgtgatgta 600
QY      121 ggtatgcacgtataaagagaagaagaataaagaatgagaataagaagaaga 180
DB      601 ggtatgcacgtataaagagaagaagaataaagaatgagaataagaagaaga 660
QY      181 cgaataataaacccaggaagaagcatttaaaagggaatcatgtaaacacattgtaaaataga 240
DB      661 cgaataataaacccaggaagaagcatttaaaagggaatcatgtaaacacattgtaaaataga 720
QY      241 gtaaaaggggaggaagcctgttaaaaaaggggcgagcgaaagcattgtagaagatgaca 300
DB      721 gtaaaaggggaggaagcctgttaaaaaaggggcgagcgaaagcattgtagaagatgaca 780
QY      301 tctgattgttttcagagatgataaagaacaattggagagaagaatatalattgttgatggtgat 360
DB      781 tctgattgttttcagagatgataaagaacaattggagagaagaatatalattgttgatggtgat 840
QY      361 attacaaaacatatattctttaagaagcattatctgtaagataagaataaaataaagaacatt 420
DB      841 attacaaaacatatattctttaagaagcattatctgtaagataagaataaaataaagaacatt 900
QY      421 tatgggaagaagatgccttatattacataaacaatttgttatgacaaagaagaatagaacc 480
DB      901 tatgggaagaagatgccttatattacataaacaatttgttatgacaaagaagaatagaacc 960
QY      481 gtaactgtataatccaatcttcggaagaatttgtatagaataactgtgaagaaggcattgaacgtt 540

```

Db 961 gtaactgtatccaaatcttcggaagatctgtcagaataacgcaaggcacggaacgct 1020
 Qy 541 tatcagaataagtagtaagaatcatalcaagggaatcttaagtaaaatcaatcaacatat 600
 Db 1021 tatcagaataagtagtaagaatcatalcaagggaatcttaagtaaaatcaatcaacatat 1080
 Qy 601 cagaataattttagatgatatataaccatataaaatgcacgtcgtacagatggacaagat 660
 Db 1081 cagaataattttagatgatatataaccatataaaatgcacgtcgtacagatggacaagat 1140
 Qy 661 ctttatcttacttaacagcttaaggacatccacagacgtcttcgttaaatcttgga 720
 Db 1141 ctttatcttacttaacagcttaaggacatccacagacgtcttcgttaaatcttgga 1200
 Qy 721 caaaatagcaatgaggtacaaagaagatcttcgaaagccttcgtacatatatacgcgca 780
 Db 1201 caaaatagcaatgaggtacaaagaagatcttcgaaagccttcgtacatatatacgcgca 1260
 Qy 781 cagcatcgtgagcttttaagcctttatcgacccggaagcctttaatcacatggataatt 840
 Db 1261 cagcatcgtgagcttttaagcctttatcgacccggaagcctttaatcacatggataatt 1320
 Qy 841 aacgacaagaataaataatcaccctcgtgagaagacttaaaagatcaacggaatgctgtcaag 900
 Db 1321 aacgacaagaataaataatcaccctcgtgagaagacttaaaagatcaacggaatgctgtcaag 1380
 Qy 901 tatgaaaaatgaggaaaaagataaaacagcatcaacactgagcgatccttatctgaa 960
 Db 1381 tatgaaaaatgaggaaaaagataaaacagcatcaacactgagcgatccttatctgaa 1440
 Qy 961 gaaggaaagagcctttaaanaagcctgcagatccatcgtgagccaaagaagaatgcaca 1020
 Db 1441 gaaggaaagagcctttaaanaagcctgcagatccatcgtgagccaaagaagaatgcaca 1500
 Qy 1021 altcatctcttaatcacaagaagaagaagccttcaaaaagaatatacaaatltgataagt 1080
 Db 1501 altcatctcttaatcacaagaagaagaagccttcaaaaagaatatacaaatltgataagt 1560
 Qy 1081 gattttttatcctacgtgagaaaaagaagctttcaaaaagctcaaaatgatatctcgat 1140
 Db 1561 gattttttatcctacgtgagaaaaagaagctttcaaaaagctcaaaatgatatctcgat 1620
 Qy 1141 tcttatcctgaagaagaagaagcctttaaataagaatcagaatgagtagtaacct 1200
 Db 1621 tcttatcctgaagaagaagaagcctttaaataagaatcagaatgagtagtaacct 1680
 Qy 1201 ttaactgaaaaaagaagaagcttttaaaaaagctgaacactgatatccaacatatgct 1260
 Db 1681 ttaactgaaaaaagaagaagcttttaaaaaagctgaacactgatatccaacatatgct 1740
 Qy 1261 ataatcaaaagctgcagaatcacaagagaggttaattgatagtccgtcaatcaacttgat 1320
 Db 1741 ataatcaaaagctgcagaatcacaagagaggttaattgatagtccgtcaatcaacttgat 1800
 Qy 1321 gtaagaagaagcagataaaaggagatctcaaaatattgagcttattatcaatcaact 1380
 Db 1801 gtaagaagaagcagataaaaggagatctcaaaatattgagcttattatcaatcaact 1860
 Qy 1361 ggaagtaacctgtacaaataaatttatgtgtatgaaataatgaaataccaacttaca 1440
 Db 1861 ggaagtaacctgtacaaataaatttatgtgtatgaaataatgaaataccaacttaca 1920
 Qy 1441 gcaacctcagtgagcttaattgataatccactgataatacaaaaatcaatgaaggtatc 1500
 Db 1921 gcaacctcagtgagcttaattgataatccactgataatacaaaaatcaatgaaggtatc 1980
 Qy 1501 ttcaatgaaatcaaaaaaattcaaaatagatattctcagtaactatartgtgtgtgt 1560
 Db 1981 ttcaatgaaatcaaaaaaattcaaaatagatattctcagtaactatartgtgtgtgt 2040
 Qy 1561 ataatagaagcctgcatalagataatgagcgttgaaatggagaatccaatcatcacca 1620

Db 2041 ataatagaagcctgcatalagataatgagcgtttgaaatggagaatccaatbatacca 2100
 Qy 1621 gatactcgagcggagatatattagaanaaaggaaagcttatataccaagaacaatcggtcgt 1680
 Db 2101 gatactcgagcggagatatattagaanaaaggaaagcttatataccaagaacaatcggtcgt 2160
 Qy 1681 gaataaagatgtacaaataatgaagcaatccgaaagaatataatgaagattgtcgt 1740
 Db 2161 gaataaagatgtacaaataatgaagcaatccgaaagaatataatgaagattgtcgt 2220
 Qy 1741 aagaatgtccaaaggtaaatagatatacaaaatccaaagcagagtaaatataaat 1800
 Db 2221 aagaatgtccaaaggtaaatagatatacaaaatccaaagcagagtaaatataaat 2280
 Qy 1801 caggaaatggaataaagcatlaaggtttacccaataatatacaaaacttatacatcaacgtg 1860
 Db 2281 caggaaatggaataaagcatlaaggtttacccaataatatacaaaacttatacatcaacgtg 2340
 Qy 1861 cataatagatatgcatccaatattgtagaagaatgcttatataatgaaatgaa 1920
 Db 2341 cataatagatatgcatccaatattgtagaagaatgcttatataatgaaatgaa 2400
 Qy 1921 aataatattcaaaagtattataaaaaaggtcaacaatcacttaagtgtatgaa 1980
 Db 2401 aataatattcaaaagtattataaaaaaggtcaacaatcacttaagtgtatgaa 2460
 Qy 1981 agatttgttttacccgatatctcctccaaatataagctgaacaataatcacatcaagat 2040
 Db 2461 agatttgttttacccgatatctcctccaaatataagctgaacaataatcacatcaagat 2520
 Qy 2041 gagatataggaacaagttatattcaaaaagggttatatggtccagaatcccggttctatata 2100
 Db 2521 gagatataggaacaagttatattcaaaaagggttatatggtccagaatcccggttctatata 2580
 Qy 2101 ctcacatgaccttcaaaaaggtgtagaatcaaggaatgagatgaggtttatacacgaa 2160
 Db 2581 ctcacatgaccttcaaaaaggtgtagaatcaaggaatgagatgaggtttatacacgaa 2640
 Qy 2161 ttgtgacatgctgtgtgatatatcgtgatatctatagataagaaccaatctgattta 2220
 Db 2641 ttgtgacatgctgtgtgatatatcgtgatatctatagataagaaccaatctgattta 2700
 Qy 2221 gttacaatcttcaaaaatctatgtatcttttaagaagaagaggttaattcaacttg 2280
 Db 2701 gttacaatcttcaaaaatctatgtatcttttaagaagaagaggttaattcaacttg 2760
 Qy 2281 tatggagaacaatgaaagcggaaatttttcgaaagcctttaggttcaatctcaag 2340
 Db 2761 tatggagaacaatgaaagcggaaatttttcgaaagcctttaggttcaatctcaag 2820
 Qy 2341 gaccatgctgaaagctttaaagttcaaaaaaatgctcgaaaaacttccaattatcaac 2400
 Db 2821 gaccatgctgaaagctttaaagttcaaaaaaatgctcgaaaaacttccaattatcaac 2880
 Qy 2401 gatcagattaaattcaatataactcataa 2430
 Db 2881 gatcagattaaattcaatataactcataa 2910

RESULT 3
 AAS00250
 ID AAS00250 standard; DNA; 1455 BP.
 XX AAS00250;
 AC
 XX
 XX
 XX
 XX
 XX
 DE LFn-Bcl-Xl apoptosis-modifying fusion protein, DNA sequence.
 XX
 XX Human; LFn-Bcl-Xl; apoptosis; cancer; spinal muscular atrophy; ds;
 KW anthrax lethal factor; neoplasm; tumour; hyper-proliferation;
 KW Alzheimer's disease; neurodegenerative disorder; stroke;
 KW transient ischaemic neuronal injury; spinal cord injury;

Query Match	31.6%;	Score 767.6;	DB 22;	Length 1455;
Best Local Similarity	99.5%;	Pred. No. 1e-124;		
Matches 770;	Conservative	0;	Mismatches 4;	Indels 0;
			Gaps	0;

[illegible]


```
XX Electron-transfer group; ETW; mismatch; genotyping;
KM gene expression; ss.
XX Synthetic.
XX WO200107665-A2.
XX 01-FEB-2001.
PD 26-JUL-2000; 2000WO-US20476.
PR 26-JUL-1999; 990US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX Umek RM;
PI WPI; 2001-159728/16.
DR Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface.
PS Example 6; Page 127; 159pp; English.
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETW) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;
SQ

Query Match 8.7%; Score 211.2; DB 22; Length 936;
Best Local Similarity 1.1%; Pred. No. 3.1e-28;
Matches 9; Conservative 559; Mismatches 222; Indels 0; Gaps 0;

OY 1182 ggtgagatagtagtaactccttctcgtacaaagaagaaggttttaaaaaagctgaact 1241
   || || : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 793 GGGGCGMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWM 734

OY 1242 tgatatcaacataatgaataatcaagaaggtgcagatacagaggttaattgtag 1301
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 733 MWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWM 674

OY 1302 tccgtcaattaatcttgatgaagaagcagataaaagsgatccaataatgtagtc 1361
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 673 MWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWM 614

OY 1362 tttaatacaatcatcgtgaagtagctgtacaaataattatctgataaataat 1421
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 613 MWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWM 554

OY 1422 gaatacaataacatacagcaacctgaagtcgagattagtgatccaactgataaac 1481
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 553 MWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWM 494

OY 1482 taaataatagaggtatttcaatgaattcaaaaaaaattcaaatatagattctag 1541
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 493 MWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWM 434

OY 1542 taactatgatgtgtatataaataagagcctcatagataatgagcgttgaatg 1601
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 433 MWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWM 374

OY 1602 gagaatcaattatcacagataactcgacagagattttagaaatggaagcttatt 1661
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 373 MWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWM 314
```

```
OY 1662 acaagaacatcgctcgtgaataaagatgtacaaataatgaacatccgaaaaa 1721
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 313 MWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWM 254

OY 1722 atataaaggatgtgcgaagtagtccaaagagtaaaatagatacaaaatccaaga 1781
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 253 MWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWM 194

OY 1782 agcacagttataataataacaggaatggaataagcatagggtaccacaaatacaca 1841
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 193 MWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWM 134

OY 1842 gcttattacatcaacgtgcataatagatagcatccaatattgtaagaagtgctatt 1901
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 133 MWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWM 74

OY 1902 aatatgaatgaatgaaataataatccaagtgatctataaaaaggtaacaatta 1961
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 73 MWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWM 14

OY 1962 cttagttgat 1971
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 13 MWMWMWMWMWM 4

RESULT 9
AAF58257/C
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
XX 24-APR-2001 (first entry)
DT
XX
DE Oligonucleotide D1954.
XX
KM Electron-transfer group; ETW; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 990US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX Umek RM;
PI WPI; 2001-159728/16.
DR Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface.
PS Example 6; Page 127; 159pp; English.
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETW) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;
SQ

Query Match 8.7%; Score 211.2; DB 22; Length 936;
Best Local Similarity 1.1%; Pred. No. 3.1e-28;
```



```
Db 253 WWWWWW.....:
Qy 1782 agcacgttaataataatcagaatggaataagcattcaggtcccaataatacaaa 1641
Db 193 WWWWWW.....:
Qy 1842 gctattacatcaacgctgacataatagatagatcccaatattgtaagaagtctattt 1901
Db 133 WWWWWW.....:
Qy 1902 aatattgaatgaatggaataataatcagaatgcttataaaaaaggtaacaatta 1961
Db 73 WWWWWW.....:
Qy 1962 cctaagtcat 1971
Db 13 WWWWWW.....4
```

RESULT 11

```
AAF58262/c
ID AAF58262 standard; DNA; 936 BP.
XX
AC AAF58262;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2007.
XX
KW Electron-transfer group; ETW; mismatch; genotyping;
KM gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 990US-0145695.
XX
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface.
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETW) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;
```

Query Match 8.7%; Score 211.2; DB 22; Length 936;

Best Local Similarity 1.1%; Pred. No. 3.1e-28; Mismatches 222; Indels 0; Gaps 0;

```
Qy 1182 ggtgagatagtaacatccttatctgaaaagaagaagttttaaaaaagctgaact 1241
11 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 793 GGGGGCGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 734
```

```
Qy 1242 tgaattcaacatagatataatcaaaagttgcaagatcacagggttaattgtag 1301
Db 733 WWWWWW.....:
Qy 1302 tccgcatcaatccttcatgtaagaagcagatataaaggatataccaatatctgagc 1361
Db 673 WWWWWW.....:
Qy 1362 ttattacatcaatccatggaagctacgttgaataaaattatttgaataaat 1421
Db 613 WWWWWW.....:
Qy 1422 gaatcaataaaccttaacagcaaccctagtgcgagattcgttgatccactgataac 1481
Db 553 WWWWWW.....:
Qy 1482 taaatlaagaggtatttcaatgattcaaaaaattcaaatatagattctag 1541
Db 493 WWWWWW.....:
Qy 1542 taactatgatgttgatataaatgaaaggcctgcatagataatgagcgttgaatg 1601
Db 433 WWWWWW.....:
Qy 1602 gagaatccaattatccagatattcgcagcagatattcgaataatggaagcttatac 1661
Db 373 WWWWWW.....:
Qy 1662 acaagaacaatcgcgtgtaataaagatgtaacaataatgaacatccgaagaaga 1721
Db 313 WWWWWW.....:
Qy 1722 atataaagatgtagcgaaagtagtgcgaagagtaaatagatacaaaaattcaga 1781
Db 253 WWWWWW.....:
Qy 1782 agcacgttaataataatcagaatggaataaagcattcaggttaccacaataatacaaa 1841
Db 193 WWWWWW.....:
Qy 1842 gctattacatcaacgctgacataatagatagatcccaatattgtaagaagtctattt 1901
Db 133 WWWWWW.....:
Qy 1902 aatattgaatgaatggaataataatcagaatgcttataaaaaaggtaacaatta 1961
Db 73 WWWWWW.....:
Qy 1962 cctaagtcat 1971
Db 13 WWWWWW.....4
```

RESULT 12

```
AAF58255/c
ID AAF58255 standard; DNA; 938 BP.
XX
AC AAF58255;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1876.
XX
KW Electron-transfer group; ETW; mismatch; genotyping;
KM gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
```

PF 26-JUL-2000; 2000WO-US20476.
XX
XX 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI UmeK RM;
XX
DR WPI: 2001-159728/16.
XX
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
XX
PS Example 6; Page 127; 159pp; English.
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX
SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query: Match 8.7%; Score 211.2; DB 22; Length 938;
Best Local Similarity 1.1%; Pred. No. 3.1e-28;
Matches 9; Conservative 559; Mismatches 222; Indels 0; Gaps 0;

QY 1182 ggtgtagtgtagtgccttctatctgaaagaaagagcttttaaaagctgaact 1241
DB 793 GGGGCGCWW 734
QY 1242 tgatattcaaccatgatgatattcaaaaggttgcaagatacagagaggttaattgatag 1301
DB 733 WWW 674
QY 1302 tccgtcaatattcttgatgaagaagcagataaaaggatattcaaatatgatgc 1361
DB 673 WWW 614
QY 1362 ttattacatcaatccatgtgaagtagcttgacaataattatgttgaataat 1421
DB 613 WWW 554
QY 1422 gaatacaataacttacagaaacctaggtcgagatttagttgacttgcactgataac 1481
DB 553 WWW 494
QY 1482 taaataatagaggtatttcaatgaattcaaaaaaattcaataatagattctag 1541
DB 493 WWW 434
QY 1542 taactatagttgttgataaataagaaagccgtcatagataatgagcgttgaatg 1601
DB 433 WWW 374
QY 1602 gagaatcaattacaccagatactcgagcaggatatttagaaatggaagctatatt 1661
DB 373 WWW 314
QY 1662 acaagaacaatcggtctggaataaagatgatacaataatgaagcaatcgaaaaaga 1721
DB 313 WWW 254
QY 1722 atataataggttgatgcgaagtagtgcgaagatgaataatagatacaaaaatcaaga 1781
DB 253 WWW 194
QY 1782 agccacgttaataataatcgaatggaataagcattaggtttaccataaatatacaaa 1841
DB 194 WWW

DB 193 WWW 134
QY 1842 gctattacattcaacgtgcataatagatgatccaatattgaaagtgcttatt 1901
DB 133 WWW 74
QY 1902 aatatgaatgaaagaaataatattcaaaagtgatcttaaaaaagtaacaaatta 1961
DB 73 WWW 14
QY 1962 cttagttgat 1971
DB 13 WWWWWWWWWWW 4

RESULT 13

AAF58252 standard; DNA; 936 BP.

AAF58252;

24-Apr-2001 (first entry)

Oligonucleotide D1835.

Electron-transfer group; ETM; mismatch; genotyping;
gene expression; ss.

Synthetic.

W0200107665-A2.

01-FEB-2001.

26-JUL-2000; 2000WO-US20476.

26-JUL-1999; 99US-0145695.

17-MAR-2000; 2000US-0190259.

(CLIN-) CLINICAL MICRO SENSORS INC.

UmeK RM;

WPI: 2001-159728/16.

Nucleic acids containing electron-transfer group, useful as labels in
hybridization assays, e.g. for genotyping, allowing repeat analyses on
a single surface

Example 6; Page 127; 159pp; English.

The present invention relates to a composition comprising two nucleic
acids each containing an electron-transfer group (ETM) having
different redox potentials. The invention is used for electronic
detection of nucleic acids, especially of substitutions (mismatches)
and single-nucleotide polymorphisms, e.g. for genotyping,
monitoring gene expression.

SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 8.6%; Score 209.4; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 6.4e-28;
Matches 6; Conservative 561; Mismatches 222; Indels 0; Gaps 0;

QY 1062 aatcaaatgatagtagtatttctatctgaggaagaaagagttttaaaagct 1121
DB 1 WWW 60
QY 1122 acaatgatattcgtgattcttctatctgaaagaaagagcttttaaatagataca 1181
DB 61 WWW 120

Oy	1182	ggtggatagatgaatcccttactcgtaaaaagaagaaggtttttaaagaagctgaact	1241
Db	121	180
Oy	1242	tgatattcaaccatgatatttaaccaaggttcgaagatcacaggagggttaattgatag	1301
Db	181	240
Oy	1302	tcgcgtcaattaatcttgaigtlaaagaagcaglataaagggaattccaatatatgatgc	1361
Db	241	300
Oy	1362	tttatccatcaatccatctggaagtcacctgtacaaataaatttcttgatgaataat	1421
Db	301	360
Oy	1422	gaatatacaataccttcacgacaacctaggtgcgagtttagttgattccacgtataatc	1481
Db	361	420
Oy	1482	taaaattaatagaggtaatttccaatgaattcaaaaaaattccaatatagattctcag	1541
Db	421	480
Oy	1542	taactatacatgattgctgataaagtgaagcgctgcattagataatgacgcttgaatg	1601
Db	481	540
Oy	1602	gagatccaattatcacccagatactcgagcagatattagaaaatggaagcttatatt	1661
Db	541	600
Oy	1662	acaaagaacatcgcgtctgtaaaataaagatgtacaataattcaagcaatccgaaaaaga	1721
Db	601	660
Oy	1722	atataataagattgtagcgaagtagtgcacaagtaaatagatacaaaaaatcaaga	1781
Db	661	720
Oy	1782	agcacagttaaataataatcgaatgcgaataaagcattagggttaaccaaaatatacaaa	1841
Db	721	780
Oy	1842	gcttattac	1850
Db	781	
RESULT 14			
ID	AAFS8254 standard; DNA; 936 BP.		
XX	AAFS8254:		
AC	AAFS8254:		
XX			
DT	24-APR-2001 (first entry)		
XX			
DE	Oligonucleotide D1875.		
XX			
KW	Electron-transfer group; ETW; mismatch; genotyping;		
KM	gene expression; ss.		
XX			
OS	Synthetic.		
PN	MO200107665-A2.		
XX			
PD	01-FEB-2001.		
XX			
PF	26-JUL-2000; 2000MO-US20476.		
XX			
PR	26-JUL-1999; 98US-0145695.		
PR	17-MAR-2000; 2000US-0190259.		
XX			

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 2, 2001, 13:49:05 ; Search time 1758.69 Seconds
(without alignments)
22794.322 Million cell updates/sec

Title: US-09-747-521-1
2430
Sequence: 1 atgataataaaagaatt.....agttcattataactcataa 2430

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_hig:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_higo_hum:*
31: em_higo_inv:*
32: em_higo_rod:*
33: em_hig_hum:*
34: em_hig_inv:*
35: em_hig_rod:*
36: em_hig_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2430	100.0	3291	1	BACIEFB	M29081 Bacillus an
2	2430	100.0	3291	6	I33395	I33395 Sequence 1
3	2430	100.0	3291	6	I69373	I69373 Sequence 1
4	2430	100.0	3631	1	BACIEFB	M30210 B.anthraxis
5	2430	100.0	181654	1	AF065404	AF065404 Bacillus
6	767.6	31.6	1455	6	AX085496	AX085496 Sequence 5
7	762.4	31.4	1368	6	I69375	I69375 Sequence 5
8	762.4	31.4	1368	6	I69375	I69375 Sequence 5
9	762.4	31.4	1425	6	I33398	I33398 Sequence 7
10	762.4	31.4	1425	6	I33398	I33398 Sequence 7
11	762.4	31.4	1524	6	I33399	I33399 Sequence 9
12	762.4	31.4	1524	6	I69377	I69377 Sequence 9
13	162.6	6.7	2980	1	BACCYA	M23179 Bacillus an
14	162.6	6.7	2980	6	A02546	A02546 B.anthraxis
15	162.6	6.7	3420	1	BACCYA	M24074 Bacillus an
16	162.6	6.7	181654	1	AF065404	AF065404 Bacillus
17	159.4	6.6	2860	1	AF031382	AF031382 Bacillus
18	158.2	6.5	2980	6	A07289	A07289 B.anthraxis
19	110	4.5	143381	9	AC091214	AC091214 Homo sapi
20	103.4	4.3	158398	2	AC011146	AC011146 Homo sapi
21	103.2	4.2	130540	2	AC079417	AC079417 Mus muscu
22	102.8	4.2	14867	3	AE001398	AE001398 Plasmodiu
23	102.2	4.2	4102	3	AF270648	AF270648 Plasmodiu
24	101.6	4.2	298469	3	AE003846	AE003846 Drosophila
25	101	4.2	2426	8	SD049822	SD049822 Saccharomyc
26	100.2	4.1	155204	2	AC007926	AC007926 Trypanoso
27	98.4	4.0	1686	22	E08995	E08995 DNA encodin
28	98.2	4.0	110000	2	PFMAL1P1_1	PFMAL1P1_1
29	97	4.0	104992	2	AC005504	AC005504 Plasmodiu
30	97	4.0	153477	2	AC006278	AC006278 Plasmodiu
31	97	4.0	169546	2	AC004157	AC004157 Plasmodiu
32	97	4.0	303091	2	AC084799	AC084799 Mus muscu
33	96.8	4.0	5420	3	AF056936	AF056936 Plasmodiu
34	96	4.0	2069	22	E10125	E10125 DNA encodin
35	96	4.0	3399	22	E10126	E10126 DNA encodin
36	95.6	3.9	67970	3	PFMAL1P3	PFMAL1P3
37	94	3.9	4995	3	PFAMESA	M69183 Plasmodium
38	93.6	3.9	54345	3	AC084152	AC084152 Caenorhab
39	93.4	3.8	47573	3	AF030694	AF030694 Plasmodiu
40	92.8	3.8	15421	3	PFCOMPIRA	X95275 P.falciparu
41	92.4	3.8	111554	8	AP002460	AP002460 Arabidops
42	91.8	3.8	213530	2	AC090493	AC090493 Mus muscu
43	91.6	3.8	7218	6	I66494	I66494 Sequence 14
44	91.2	3.8	120029	2	HSJ282H10	AL132672 Homo sapi
45	90.4	3.7	183584	9	AC012492	AC012492 Homo sapi

ALIGNMENTS

RESULT	1	LOCUS	BACIEFB	3291 bp	DNA	BCT	12-OCT-1995
DEFINITION		Bacillus anthracis			lethal factor (lef) gene, complete cds.		
ACCESSION		M29081					
VERSION		M29081.1		GI:143143			
KEYWORDS		anthrax toxin; exotoxin.					
SOURCE		Bacillus anthracis (clone: pLF74) DNA.					
ORGANISM		Bacillus anthracis					
REFERENCE		Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Bacillus; Bacillus cereus group.					
AUTHORS		Brayg,T.S. and Robertson,D.L.					
TITLE		1 (bases 1 to 3291)					
JOURNAL		Nucleotide sequence and analysis of the lethal factor gene (lef)					
MEDLINE		from Bacillus anthracis					
FEATURES		Gene 81 (1), 45-54 (1989)					
		90034185					
		location/Qualifiers					
		1..3291					
		/organism="Bacillus anthracis"					

QY 1681 gaaataaagatgtacaataaataaagcaatccgaaaaagaatataaagatgtgcg 1740
|||||
Db 2161 GAAATTAAGATGTACAAATTAATTAAGCAATCGAAAGAAATATTAAGATGTGATCG 2220
QY 1741 aaagtgtgccaaagataaataagatacaaaaatcgaagacagatlaataaataat 1800
|||||
Db 2221 AAAGTGTGCCAAAGATTAATTAATTAAGCAAAATTAAGCAAAATTAATTAATTAAT 2280
QY 1801 caggaaatgaataaagcattaggtttacaaaatatcaaaaagcttaattcattcaacg 1860
|||||
Db 2281 CAGGAATGAATTAAGCAATTAAGGTTACCAAAATTAACAAAGCTTATTAACATTAACG 2340
QY 1861 cataatagataatgcataatcattagaaagtgtcatttaataatgaatgaatga 1920
|||||
Db 2341 CATTAATAGATGTACCAATTAATTAAGCAAAAGCTTATTAATTAATTAATTAATGA 2400
QY 1921 aataatattcaaaagtgtattcaaaaagaatacaaatcttagttagtgatga 1980
|||||
Db 2401 AATTAATTAATTAAGCAATTAATTAAGCAAAAGCTTATTAATTAATTAATTAATGA 2460
QY 1981 aaattgttttaacgataatctctcctaataatagctgaacaataacataaagat 2040
|||||
Db 2461 AATTTGTTTACCGATTAATTAATTAATTAAGCAAAAGCTTATTAATTAATTAATGA 2520
QY 2041 gaaatataagcaagatcattcaaaaagggttataatcctcgaacatccgcttata 2100
|||||
Db 2521 GAGATTAATGAACCAATTAATTAAGCAAAAGGTTATTAATTAATTAATTAATTAATGA 2580
QY 2101 ctccatgtgacctcaaaaggtgtagaataaagatgaatgaatgaatgaatgaatga 2160
|||||
Db 2581 CTCCAATGAACCTTCAAAAGGTTAGAAATTAAGCAATTAATTAATTAATTAATTAATGA 2640
QY 2161 tttagacatgctgtgataatataatgctgataatcattagataaagaacaatcgttta 2220
|||||
Db 2641 TTTGGACATGCTGTGATTAATTAATTAATTAAGCAATTAATTAATTAATTAATGA 2700
QY 2221 gttaacaattcnaaaatcattatgataatcatttaagaagaagaagatlaattcaatc 2280
|||||
Db 2701 GTTACAAATTTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 2760
QY 2281 taatggagaacaataagcgaatcttttcgaagaaccttaggttaatgcatctcag 2340
|||||
Db 2761 TATGGGAGAAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 2820
QY 2341 gaccatgctgaacgtttaaaagttcaaaaagcctcgaagaaccttcaattatatac 2400
|||||
Db 2821 GACCATGCTGAACGTTTAAAGTTCAAAAATTAATTAATTAATTAATTAATTAATGA 2880
QY 2401 gatcagatlaagttcattatlaactcataa 2430
|||||
Db 2881 GATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 2910

RESULT 2
LOCUS 13395 3291 bp DNA PAT 06-FEB-1997
DEFINITION Sequence 1 from patent US 5591631.
ACCESSION 13395
VERSION 13395.1 GI:1824186
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 3291)
AUTHORS Leppla,S.H., Klumpel,K.R., Arora,N., Singh,Y. and Nicholls,P.J.
TITLE Anthrax toxin fusion proteins, nucleic acid encoding same
JOURNAL Patent: US 5591631-A 1 07-JAN-1997;
FEATURES Location/Qualifiers
source 1..3291
BASE COUNT 1333 a 391 c 564 g 1003 t
ORIGIN

Query Match 100.0%; Score 2430; DB 6; Length 3291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgaataaataaagaattataaagaataatagatagatgcatgtttagtaacagcaat 60
Db ATGAATATAAATAAAGAAATTAATTAAGCAATTAATTAATTAATTAATTAATTAATGA 540
QY 61 actttagtggtccgtcttattcccccctgtcaggggcyggtcagtgatgta 120
Db 541 ACTTTAGTGTCCCGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 600
QY 121 ggtatcagcgttaaaagaaagaaagaaataaagaatgagataaagaaagaaagaa 180
Db 601 GGTATGCGCTTAAGGAAAGAAAGAAATTAAGCAATTAATTAATTAATTAATTAATGA 660
QY 181 cgaataaataaagcagagagcatttaagaataatcagaacacattgtaaaataga 240
Db 661 CGAATATAAACAAGAGAGCATTTAAAGCAATCATGAACACATTTGTAATAATTAAG 720
QY 241 gtaaaaggaggagagcgtgttaaaagaagcagcaaaagcacttctgaagaatca 300
Db 721 GTAAAGGGGAGAGAGCTTTAAAGCAAGGCAAGAAAGCTTGTGAAGAAAGTAACA 780
QY 301 tctgatgttttagagatgtaataagaatggaagaaagaaatataatgtgtgagat 360
Db 781 TCTGATGTTTTAGAGATGTAATTAAGCAATTAAGCAAGGAAAGATTAATTAATTAATGA 840
QY 361 attcaaaacataatcatttagaagcattcgtgaagataaagaataaataaagaacat 420
Db 841 ATTACAAACATTAATTAATTAAGCAATTAATTAATTAATTAATTAATTAATTAATGA 900
QY 421 taagggaagaatgcttattataatcagaaatcattatgataatcagaagaatcgaacc 480
Db 901 TATGGGAAAGATCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 960
QY 481 gtaactgttaacaaatcctcgaagaatgataatgaataaatacgaagaagcgaacgt 540
Db 961 GTACTTGTAAATCCAAATCTTGGAGAGATTATGTAAGAAATTAATTAATTAATTAATGA 1020
QY 541 taatgtgaataagtagaataatatacaaggagataatttaagtaaaatlaaacaacat 600
Db 1021 TATTATGAATTAAGTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 1080
QY 601 cagaataatcattagatgataatataacataaagaatcagatcagatgagaagaat 660
Db 1081 CAGAAATTTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 1140
QY 661 ctttatttaataatagccttaaggaaacatcccaagacattctgtagaattctggaa 720
Db 1141 CTTTATTATTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 1200
QY 721 caaaatagcaatgaggtacagaagatattgtcgaagaaccttttcataatataatcag 780
Db 1201 CAATAATTAAGCAATGAGGAGACAAAGATTTTGGCAAGGCTTTGCAATTAATTAATGA 1260
QY 781 cagcatcgatgagtttataagccttatgcaacggaagccttataatataatgataatc 840
Db 1261 CAGCATCGTATGATTTTATGAGCTTTATGACACCGAAGCCTTTTATTAATTAATTAATGA 1320
QY 841 aacgaacaaagaataaataatcatalccttggaaagaaacttaagaatacaagcagctga 900
Db 1321 AACGAACAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 1380
QY 901 tatgaataaatgggaaagaataaagaacagcactatacaacacggaagcagattcattcga 960
Db 1381 TATGAATAAATGGGAAAGATTAATAAGCACTATCAACACGTGAGGCAATTTTATTAATGA 1440
QY 961 gaaagaaagagactttaaagaagcgtcagatctcattatgagccaaagaagaatgacata 1020
Db 1441 GAAAGAAAGAGACTTTTAAAGAGCTGACAGATCTTAATTAATTAATTAATTAATTAATGA 1500

```

OY 1021 atcaatcttctatctcaagaagaaagagcttcaaaagaacatacaaatgtacagta 1080
    |||||||
Db 1501 ATTCAATCTTATCTCAAGAAAGAAAGAGCTTCAAAAAGAAATACAAATGATAGTAGT 1560
OY 1081 gattttctactcgaagaaagagctttaaanaagctacaaatgtatctcgtgat 1140
    |||||||
Db 1561 GATTCTTAACTACTGAGGAAAAAGACTTTTAAAAAGCTACAAATGTGATATCTCGTAT 1620
OY 1141 tctctctcgaagaagaaagagctttaaataagaatacaggtggaagtagtaacct 1200
    |||||||
Db 1621 TCTTTTCTGAAAGAAAAAGAGCTTTTAAATAGAAATACAGGTGAGTAGTAAATCC 1680
OY 1201 ttaactcgaagaagaaagagctttaaanaagctgaactgtatctcaacatactga 1260
    |||||||
Db 1681 TTATCTGAAAGAAAAAGAGCTTTTAAAAAGCTGAACTGATATTCACCATATGAT 1740
OY 1261 atcaatcaaggtctgcaagatacaggaaggttaattgtacagtcgcgaacttaactgat 1320
    |||||||
Db 1741 ATTTAATCAAAAGGTGCAAGATACAGAGGCTTAATTGATAGTCGCTCAATTAATCTTGAT 1800
OY 1321 gtaagaagcagatacaaaagagatctcaaaatctgactcttaatacatcaatccatt 1380
    |||||||
Db 1801 GTAAAGAAAGCACTATPAAAGGATATTCAAATATTCAGTTTATTCATTAATCCAT 1860
OY 1381 ggaagtaacctgtacaataaaattattgtatgaaataatgaatatacaataaccttaca 1440
    |||||||
Db 1861 GGAAGTACCTGTGTACAAATTAATTTATTGTATGAAATATGAAATATCAATTAACCTTACA 1920
OY 1441 gcaacctcaggtctggaattgtatgtatccactgaataactaaataaagagattc 1500
    |||||||
Db 1921 GCAACCCCTAGCTCGGATTTAGTTGATTCACCTGATTAATTAATAATTAAGAGTAT 1980
OY 1501 ttaactgaattcaaaaaaaattccaataatagatcttctagtaactaatgattgattgat 1560
    |||||||
Db 1981 TTCAATTAATTCAAAAAAATTTCAAAATATGATTTCTAGTAACTATATGATTTGTTGAT 2040
OY 1561 ataaatgaagagcctgcatctagatacagtcggttgaaatggagaaatccaattaccca 1620
    |||||||
Db 2041 ATTAATGAAGGCTGCTGATTAATGATTAATGACGTTTGAAGGAAATCCAAATTAATACCA 2100
OY 1621 gatactcgaagcagatattagaanaatggaagccttaatacaagaagaacatcgctcgt 1680
    |||||||
Db 2101 GATACCTGACAGGATATTAGAAATGGAAGCTTAATTAATAAGAAACATCGGCTCG 2160
OY 1681 gaaataaagagctgtacaataaatttaagcaatccgaaagaagataataagattgtacg 1740
    |||||||
Db 2161 GAAATTAAGAGTACAAATTAATTAAGCAATCCGAAAAAGAAATATTAAGGATTAATGCG 2220
OY 1741 aaagttagtccaaagataaatacagaanaatcaagaagcagatgaataataat 1800
    |||||||
Db 2221 AAAGTAGTCCCAAGAGTAAATAGATTAATAAATTCAGAAAGCACAGTTAATATTAAT 2280
OY 1801 caggaatggaataaagcattaggtctacaaataatacaagccttaatacaagcg 1860
    |||||||
Db 2281 CAGGAATGGAATTAAGCAATTAAGGTTACCAAAATATACAAACCTTAATTAATCAATTAAC 2340
OY 1861 cataatagatactcgaataatgttaagaagctcttaataatctgaatgaatgaa 1920
    |||||||
Db 2341 CATATATATATGCTATCCAAATTTGTAGAAAGTCTTAATTAATTAATTAATTAATTAAT 2400
OY 1921 aataatattcaagtgactctataaagaagtaacaattcttaagtgatgaa 1980
    |||||||
Db 2401 AATATATATTCAAAGTGATCTTAATTAATAAAGGTAAACAAATTAATTAATTAATTAAT 2460
OY 1981 agattttttttacagatattactctccataataagctgaagaatatacaacaaagat 2040
    |||||||
Db 2461 AGATTCTTTTAAACGATATTAATCTCCCTAATATTAATTAATTAATTAATTAATTAAT 2520
OY 2041 gagaatagaagaagcttcaataaaggtatatagtctcagaatccgttctata 2100
    |||||||
Db 2521 GAATATATAGCAAGTTCATTCAAAAAGGTTATATGTCTCCAGATCCGTTCTATATTA 2580

```

```

OY 2101 ctccatgaccttcaaaaagtgtagaattaaaggaatagatgtgaggttctatacagaa 2160
    |||||||
Db 2581 CTCATGAGACCTTCAAAAAGGTGATTAATTAAGCAATATATGAGGCTTTATTAACAGAA 2640
OY 2161 ttggagaatcgtgtgagatctatctgtggaatcatalatagaataagaacaatctgatt 2220
    |||||||
Db 2641 TTTGGACATCTGTGATGATTAATGCTGATATCTAATTAATTAAGAACCAATGTGATTA 2700
OY 2221 gtacaaatcttaaaaaaatcatctgatatctttaaaggaagaaggaatcaatctcg 2280
    |||||||
Db 2701 GTTACAAATTTCTAAAAAATTCATTTGATTAATTTTAAAGAAAGAGGAGTAAATTAATTA 2760
OY 2281 tatggagaagaacaatgaagcgaattcttgcagaagccttaagtgatctacag 2340
    |||||||
Db 2761 TATGGAGAAACAATGAACCGGATTTTTCAGAAAGCTTTAATGATTAATGATTAATTA 2820
OY 2341 ggcctatgcgaacgtttaaagttcaaaaaaattgctcgaagaacttccaattatcac 2400
    |||||||
Db 2821 GACCATGCTGAACGTTTAAAGTCAAAAAAATGCTCCGAAACCTTCCAAATTTATTAAC 2880
OY 2401 gatacagattaagttcatatcatctacataa 2430
    |||||||
Db 2881 GATCAGATTAAGTTCATTAATTAATTAATTA 2910

RESULT 3
LOCUS 169373 3291 bp DNA
DEFINITION Sequence 1 from patent US 5677274.
ACCESSION 169373
VERSION 169373.1 GI:2831495
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 3291)
AUTHORS Leppla,S.H., Klimpel,K.R., Arora,N., Singh,Y. and Nichols,P.J.
TITLE Anthrax toxin fusion proteins and related methods
JOURNAL Patent: US 5677274-A 1 14-OCT-1997;
FEATURES
source 1..3291
BASE COUNT 1333 a 791 c 564 g 1003 t
ORIGIN

Query Match 100.0%; Score 2430; DB 6; Length 3291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgaatataaaaaaagaatttaaaagaatagatagtcgtttgtaacagcaatc 60
    |||||||
Db 481 ATGAATATPAAAAAAGAAATTTTAAAGATTAATGATGTCATGTTTGTAAACACCAAT 540
OY 61 acttgagtggtcccgctcttataccctgtacaggggagggcggtcagtgatgta 120
    |||||||
Db 541 ACTTTGAGTGTGTCCTGCTTTATCCCTTGTACAGGGGCGGCGGTCAATGATGATGTA 600
OY 121 gttatgcagcttaaaagagaagaagaataaagaatgaatgaagaagaagaaga 180
    |||||||
Db 601 GGTATGCACTGTAAGAAAGAAAGAAATAAAGATGAGATTAAGAAAGAAAGAAAGAA 660
OY 181 cgaataaaacacagagagcatttaaggaatctgaacaacatctgtaaaatgaa 240
    |||||||
Db 661 CGAATTAACACAGAGAAAGACATTTTAAAGGAATTCATGAAMCATTGTAAATAATGAA 720
OY 241 gtaaaaggaggaagagctgttaaaagaagcgaacgaagaagctacttggaagta 300
    |||||||
Db 721 GTAAAAAGGAGAGAGCTGTTAAAAAAGAGCAGCAAAAAGCTACTTGAGAAAGTACCA 780
OY 301 tctgagttttaagatagataaagcaatcggaggaagatatatactgtgagtgat 360
    |||||||
Db 781 TCTGATGTTTAAAGATGATTAAGCAATTCGAGGAAAGATATATATTGTGATGCTGAT 840

```

OY 361 attcaaaacataatcctttagaagcattatcgaagataagaanaaaataaagacatt 420
|||||
Db 841 ATTACAAAACATATATCTTTAGAGCATATCTGAAGATATAGAAAATAAAGACATT 900
OY 421 tatgggaagatgcttattacatgacaatlatglatatgcaaaagaagatagaacc 480
901 TATGGAAAGATGCTTATTTACATGAACATTATGTATATGCAAAAAGAGATATGAACCC 960
OY 481 gtaactgtataccaatcttcggaagaattatgtagaanaataactgaaagcactggaacgt 540
|||||
Db 961 GATACCTGTATCCAACTTCGGAAGATATGTAGAAAATATGAAAAGCAGCAACGTT 1020
OY 541 tattatgaataatgtaagaatattatcaaggatattlaagtaaaaattaatcaacat 600
|||||
Db 1021 TATTATGAAATAGTAAAGTATTTATCAAGAGATTTTAAAGTAAATTTAATCAACATAT 1080
OY 601 cagaatattttagatglatataaataccatlaaanaalgcaltgaltcagatggaacaagat 660
|||||
Db 1081 CAGAAATTTTATGATGTATTAATACCATTAATAAATGCATCTGATTCAGATGACAAGAT 1140
OY 661 ctttatttaccatcagcttaaggaacatccacagactttctgtagaattcttggaa 720
|||||
Db 1141 CTTTATTTACTAATCAGCTTAAGGAACATCCACAGACTTTTCTGTAGATCTTGAA 1200
OY 721 caaaatagcaatgaggtacagaagaatatttgcgaagaactttgcatattatatacgaacca 780
|||||
Db 1201 CAAATATGCAATGAGGTATCAAGAGATTTTGCAGAAAGCTTTTGCAATATATATGAGCCA 1260
OY 781 cagcatcgtatgcttctacagcttctacgacggaagcttlaatacatgataaatt 840
1261 CAGCATCGATGATTTTACAGCTTTATGCACCGAGCTTTTAAATATACATGATTAATTT 1320
OY 841 aacggaacaaataaatacctatcccttggagaacttaagaatcaacgagatgctgcaaga 900
|||||
Db 1321 AACGAACAAGAAATATATCTATCTTGAAGAACTTAAAGATCAAGATGCTGTCAAGA 1380
OY 901 tatgaaaaatgggaaaaagataaaacagcaatcaaacactcggagcattcttattatga 960
|||||
Db 1381 TATGAAAATGGGAAAAGATTAANAACAGCACTATCAACACGAGCAATCTTTATCTGAA 1440
OY 961 gaaggaagaagacttctaaanaagctgcagatctcattctgaagcaagaagaatgacata 1020
|||||
Db 1441 GAAGGAAGAGACTTTTAAAAAGCTGCAGATTCCTATTGAGCCAAAAGAAAGATGACAT 1500
OY 1021 attcattcttattcctaagaagaagaagccttcaaaaagaatacaaatltagtagt 1080
|||||
Db 1501 ATTCAATCTTTATCTCAGAGAGAAAAGAGCTTTAAAAAGAAATACAAATTTGATAGAT 1560
OY 1081 gatttttctacigaggaagaagaagctttttaaanaagctacaattltagtagt 1140
|||||
Db 1561 GATTTTCTTATCTAGAGAAAAGAGTTTAAAAAGCTTACAAATTTGATTTGCGAT 1620
OY 1141 tcttctacgaaagaagaagaagccttctaataagatacaggtggaatagtaactcct 1200
|||||
Db 1621 TCTTTATCTGAGAGAGAAAAGAGCTTTTAAATAGAAATACAGTGTATATATATCT 1680
OY 1201 tcatctgaaagaagaagaagcttctaaanaagctgaaactltagatctcaacataatg 1260
|||||
Db 1681 TATATCGAAAAAGAAAAGAGTTTAAAAAGCTTGAACCTTATTTCAACCATATGAT 1740
OY 1261 attaatcaaaaggttgcagaatagacagaggttataatgtagtccgcaatlaattctgat 1320
|||||
Db 1741 ATTAAATCAAAAGGTTGCAAGATACAGAGGGTTAATTGATATGCTGCAATTTAATCTTGAT 1800
OY 1321 gtaagaagaacagataaagaagatctcaaaaatcagatgagcttattatacaatcatt 1380
|||||
Db 1801 GTAAGAAAACAGATATTAAGGGATATTCAAAATATGATGCTTATATACATCATTCATT 1860
OY 1381 ggaagtagcttgcataaataatttctgataaanaataatgaataatcaataacttaca 1440
|||||
Db 1861 GGAAATGACTCTGTACAAATTAATTTATTTGTATGAATAATTTGAATATCAATTAACCTTACA 1920

OY 1441 gcaacctagtgcggaattagttgatttccactgataataactaaataaagaattat 1500
|||||
Db 1921 GCAACCTTAGTCCGGATTTAGTTGANTCCATGATATATCTAAATTTATATAGAGATT 1980
OY 1501 tccaatgaaltcaaaaaaaattccaatatagtaattcttagtaactatagattgtagt 1560
|||||
Db 1981 TTCATGCAATTCAAAAAAATTTCAAAATATAGTATTTCTGTATCATATATGATTTGAT 2040
OY 1561 ataaatgaagagcctgcatagataatgagcgtttgaagaatggaatccatataacca 1620
|||||
Db 2041 ATAAATGAAGGCTGCTGATTAGATTAATGAGCGTTTGAAATGGAGAAATCCAAATTAACCA 2100
OY 1621 gatactcgaacagagataatttagaanaatggaagccttataataagaanaaacatcgctg 1680
|||||
Db 2101 GATACTCGACAGATATTTAGAAAATGGAACCTTATATACAAAGAAACATCGGCTG 2160
OY 1681 gaaataaaggatgtacaaataatgaacatccgaaanaagaatataaagattgtagc 1740
|||||
Db 2161 GAATTAAGAGATGTACAAATTAATTAAGCAATCCGAAAAAGAAATATTAAGATTTGATGCG 2220
OY 1741 aaagtatgccaagaagtaaaatagatagacaaaatccaagaacacaggttaaatataat 1800
|||||
Db 2221 AAAGTATGCCAAAGAGATAAATAGATACAAAATTCAGAACACAGCTTAATATTAAT 2280
OY 1801 caggaatggaataaagcattaggttaccaaaatatacaagcttattacatcaacgtg 1860
|||||
Db 2281 CAGGATGGAATTAAGCATTTAGGCTTACCAAAATATACAAAGCTTATACATTCACAGCTG 2340
OY 1861 cataatagatatgcatccaatattgtagaagaatgcttatttaalatgaaatgaatgga 1920
|||||
Db 2341 CATATATAGATATCATATCCATATTTGTAGAAAGTGCTTATTAATATGATGAATGAGAA 2400
OY 1921 aataatattcaaaatgacttataaanaaggtaaacaaatctactagttgtagtgaatga 1980
|||||
Db 2401 AATTAATATTCAAAAGTATCTTAATAAANAAGTAAACAATTTACTAGTTGATGATTAAGGA 2460
OY 1981 agattgttttctacgatatctctcccaatagcgtgaacaaatatacaacatcaagat 2040
|||||
Db 2461 AGATTTGTTTTTACCGATATTAATCTCTCCTAATATATAGCTTAACAAATATACATCAAGAT 2520
OY 2041 gagatatatgagcaagtcttcaaaaagggttatalgttccagaatccgcttccataat 2100
|||||
Db 2521 GAGATATATAGCAAGATTCATTCAAAAGGCTTATATGCTCCAGAAATCCGTTCTATATTA 2580
OY 2101 ctccaatgaccttcaaaagggtgtagaataatgaagaatgagatgtaggggttctatacagaa 2160
|||||
Db 2581 CTCATATGACCTTCAAAAGGCTGTAAGAAATTAAGAAATGATAGTGAAGGCTTTTAAACAGAA 2640
OY 2161 ttggagacatgctgtggaatgattatgctggaatctcatatagataagaacaaatcgatct 2220
|||||
Db 2641 TTTGGACATCTGTGATGATTTATGCTGATATCTATTAAGTAAGAACATCTGATTTTA 2700
OY 2221 gttaaaaattctaaanaatcattatgatatatttlaaggaagaaggagtaattcaactcgt 2280
|||||
Db 2701 GTTACAAATTTCTAAAAAATTCATTTGATTAATTTTAAGAGAAAGAGATTAATTAACCTTCG 2760
OY 2281 tatgggaagaacaaatgaagcggaatttttgcgaagcctttagtgaatgacatctcaag 2340
|||||
Db 2761 TATGGGAACAACAAATGTAAGCGGAATTTTTCAGAAACCTTTAGTTAAGCATATTCACG 2820
OY 2341 gaccatgctgaagctttaaagaatcacaanaaaatgctccgaaacacttccaattatgaac 2400
|||||
Db 2821 GACCATGCTGACGTTTAAAGTTCAAAAATAATGCTCCGAAAACCTTCCAAATTTATTAAC 2880
OY 2401 gatcagaatlaagtcattatataactcaaa 2430
|||||
Db 2881 GATCAGATTAAGTTCATATTAACCTCAATA 2910

RESULT 4
BACLEF BACLEF 3631 bp DNA BCT 26-APR-1993
LOCUS B.anthraxis plasmid pX01 lethal factor (lef) gene, complete cds.
DEFINITION

ACCESSION M30210
 VERSION M30210.1 GI:143141
 KEYWORDS lethal factor.
 SOURCE B.anthraxis (strain Sterne) plasmid pX01 DNA, clones pLF7 and pBKSPF.
 ORGANISM Bacillus anthracis
 Bacteria, Firmicutes; Bacillus/Clostridium group;
 Bacillus/staphylococcus group; Bacillus; Bacillus cereus group.
 LOWE,J.
 REFERENCE 1 (bases 1 to 3631)
 AUTHORS A comparison of Bacillus anthracis sequences
 TITLE Unpublished
 JOURNAL Draft entry and computer-readable sequence for [1] kindly submitted
 COMMENT by J.Lowe, 29-NOV-1989.
 FEATURES
 source
 1..3631
 /organism="Bacillus anthracis"
 /db_xref="taxon:1392"
 misc_feature
 585..592
 /note="044 region"
 -35_signal
 594..599
 -10_signal
 620..625
 mRNA
 633..>3631
 /product="lethal factor mRNA"
 RBS
 669..675
 /note="ribosome binding site"
 sig_peptide
 685..783
 /note="lethal factor signal peptide"
 CDS
 685..3114
 /note="lethal factor precursor"
 /transl_start=1
 /transl_table=1
 /protein_id="AA22569.1"
 /db_xref="GI:143142"
 /translation="MNKKKFKIVISMSCLVATITLSPGVPLPVLVAGAGHGVDGMH
 KKKEKKDKENKRDENKTOEBHLEIKMHIKIVKEEAVKKEAAKLEKVED
 VLEMYATGKTIYDGTIKHLSLESEDKKIDIGKALDHEHYVKAEGEP
 VLTOSSEDEVETEKALNVEYIEGKILSRDLISKINOPYKFLDVLNTIKNASDSG
 ODULTNOLKEHPDFSVFLEQNSNEVOEFAKAPAYIIEPOHROVLOLYAPEAN
 MKPFNEOELNLSLELKIDRMISRYKKEKIKOHQWMSDSISEBGRLLKLIQITE
 PKDDIHSLSOEKELKRIODSDSPSTEBKEPLKLIQIDIRDSLESEKELNRO
 IOVDSNPSEKKEPKLKLIDIOPYDINORLIDOGGLIDSPSINLDVROKRRDIO
 NIDALLHOSIGSTLTKITLYLENMNNITATIDVSTNTKINRIGIEPKNE
 KITSNNMIVDINERPALDNERLKMRIOLSPDTRAGYENGLILORNIETIDOE
 IKOSEKEVIRIDAKVPSKIDTKIOEOLNINQEMNKALCLPYTKITITNHRV
 ASNIVESATLIDENKNNISDLIKKVTNYLDNGREFPTDITLPTNAEOYTHODEI
 YEOVSHKGLYVPSRSILHGPSKVELANDSEGRIFHEGHAVDDYAGVILDKNSDL
 VTNSKFPDIFKEGSSNLTSTGRINEAFEAFLMSTDAERLKVKQKNAPKTFQF
 INDOI.KFTINS"
 mat_peptide
 784..3111
 /note="lethal factor"
 repeat_region
 1990..2064
 /note="inverted repeat"
 repeat_region
 1990..2067
 /note="inverted repeat"
 misc_signal
 3334..3529
 /note="inverted repeat"
 repeat_region
 3334..3436
 /note="transcription terminator"
 repeat_region
 3351..3429
 /note="inverted repeat"
 repeat_region
 3351..3429
 /note="inverted repeat"
 misc_feature
 3358..3524
 /note="cruciform"
 BASE COUNT 1455 a 447 c 602 g 1127 t
 ORIGIN 22.3 kb in the 28.4 kb BamHI B fragment.
 Query Match 100.0%; Score 2430; DB 1; Length 3631;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 atgaataaaagattataaagtagtgcgtatgcatgtttagtaacagcaatt 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 685 ATGAATATATAAAGAAATTTATATAAGTAATAGTATGTCATGTTAGTAACAGCAAT 744
 QY 61 actttagatggtccgcgtcttataccccccttgcacagggcgggcggtcatggtatgta 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 745 ACTTTGAGTGGTCCGCTCTTTATCCCTTGTACAGGGGGCGGCGTCATGATGTGA 804
 QY 121 ggtatgcacgtaaagagaagaagaagaagaagaagaagaagaagaagaagaaga 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 805 GGTATGACGTAAAGAGAGAAAGAAAATTAAGATGAGAAATGAAGAAATGACAGAA 864
 QY 181 cgaataaaacacaggaagagcatctaaagaaatcagaacacacattgtataaataaga 240
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 865 CGAAATATAAACACAGAGAGACATTTAAAGAAATCATGAATACATTTGTAAATATAGA 924
 QY 241 gtaaaagggagggaagcgtgttaaaaaagagcagacagaagaactcttgagaagtacca 300
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 925 GTAAAGGGAGGAGGAGCGTTTAAAAAGAGCGACAGAAAGCTCTTGAGAAAGTACCA 984
 QY 301 tctgattcttagagatgtataaagcaattggaggaagaagatatatctgtgagtgatgac 360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 985 TCTGATGTTTATGAGATGTATTAAGCAATGTGAGGAAAGATATATGTGTGATGATGAT 1044
 QY 361 attcaaaacatatatctttagaagcatatctgaagataagataaagaagaagaatcatt 420
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1045 ATTCAAAACATATATCTTTAGAGCATATATCTGAGATAGAAAATTAAGAACATTT 1104
 QY 421 tatggagaagtgctttatatacagaaatcagaaatcagaaatcagaaatcagaaatcag 480
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1105 TATGGAAAGATGCTTTATTAATACATGAACATTAATATGCAAAAGAGATATGACACC 1164
 QY 481 gtactgttaaccacatctcgaagatcagaaatcagaaatcagaaatcagaaatcagaaatc 540
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1165 GTACTTGTATTCATCTTGGAAGATTTATGTAGAAAATCTGAAAAGGCACTGAAAGTT 1224
 QY 541 tattatgaataagtgatagatatacgaaggatatttaagtaaaatcaaccacat 600
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1225 TATTATGAATATAGTATGATATATCAAGGATATTTAGTAAATTAATCAACCATAT 1284
 QY 601 cagaatctttagatgattataaatacacaataaataagcatcgtatcagatggaagaat 660
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1285 CAAATATTTTATATGATATTAATACCATTAATAAATGCAATCTGATGAGTGAACAAAT 1344
 QY 661 ctttatttactaatacagcttaaggaacatccacagactcttcttagaatcttgaga 720
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1345 CTTTTATTTACTAATACAGCTTAAGAAACATCCACAGACTTTTCTGTAGAAATCTTGGA 1404
 QY 721 caaaataagcaatgaggtacagaagatctgcgaagctttgcatatatacagacca 780
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1405 CAAATATGCAATGAGGTACAGAAATTTGCGAAGCTTTTGCAATATATCGAGCCA 1464
 QY 781 cagcatcgatgattttttagaagctttatgacacgggaagctttaaatacgtgataatt 840
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1465 CAGCATCGATGATGTTTACAGCTTTATGACCGGAGACTTTTAATTAATGATTAATTT 1524
 QY 841 aacgaacaagaataaataatcctcttgagaagacttaagaatcaacggatgctgcaaga 900
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1525 AACGAACAAGAAATTAATCTATCTTGAGAGAACTTAAGATCAACGATGCTGCAAGA 1584
 QY 901 tatgaataaattggaaagaataaacaagcaatacaacatgagagatcttatacgtgaa 960
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1585 TATGAAAATGGAATAATTAAMACGACATTAACACTGAGGAGGATTTTATCTGAA 1644
 QY 961 gaaggaagagagactttaaagaagctgcaagttcctatcttgagccaaagaagaatgacata 1020
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1645 GAAGGAAGAGGACTTTTAAAAAGCTGAGATCTCTATTAAGCCAAAGAAAGATGACATA 1704
 QY 1021 attaatcttatacacaagaagaagaagactttaaagaatacaaatgataagtagt 1080
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1705 ATTCATCTTATCTCAAGAGAAAGAGAGCTTCTAATAAATATCAATATGATAGTAGT 1764
 QY 1081 gatttttattctactgggaaagaagagttttaaagaagctacaatgataatcgtgac 1140
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1765 GATTTTATCTACTGAGGAAAGAGAGTTTAAAAAGCTTACAAATTTGATATTCGAT 1824

Oy	1141	tcctactctgaagaagaagaaggcctttaaataagaacagtgagatgaatcct	1200
Db	1825	tctttatctgacagacagaaaaagccttttaaatgaaatcacagtgatgaaatcct	1884
Oy	1201	ttactcgaagaagaagaaggcttctttaaagaagctbaaactgcatatcaacatagat	1260
Db	1885	ttatctcgaaaaaagaaaaaggttttttaaaaaagctgaaccttgatattcaacatgat	1944
Oy	1261	atatacaaaaggttcgaagatacagaagaaggttaattgataagtcgtaacaaatcttgat	1320
Db	1945	atttaattaaaaggttcgaagaatpacagagaggttaattgaaatgacccctcaattattcttgat	2004
Oy	1321	gtagaagaagcagatataaaagggatctcaaaaatatgtagtgccttatcaatcaactt	1380
Db	2005	gtaaagaagaagcagatataaaagggatctcaaaaatatgtagtgcctttttacatcaatccatt	2064
Oy	1441	ggaaagccttaagtcggaattagttgattccactgtaaacacaaataataagagttat	1500
Db	2125	gcaaaccttagctgcggaattagttgattgattgattccactgtaaacacaaataataagagttat	2184
Oy	1501	ttcaatgaatcacaagaanaaatttcaaaataagtaattctcagtaaacataatgattgat	1560
Db	2185	tttcaatgaattcacaagaanaaatttcaaaatpagaattttgtagtaactatgatgattgattgat	2244
Oy	1561	ataaataagaagccttgcatatgaataatgagcgtttgaaatgagaatccaattacacca	1620
Db	2245	atttaattaaaagccctgcattgaaatgaaatgagcgtttgaaatgagaatccaattatcacca	2304
Oy	1621	gatactcgacagataatttagaagaatggaaagcctatacttcaagaagaacatcgctgtc	1680
Db	2305	gattctcgacagagatatttagaagaatggaaagcctattatttcaaaagaacatcgctgtc	2364
Oy	1681	gaaaataaaggatgtaacaataaattaaagcaatccgaaaaaatatataagaattgtagcg	1740
Db	2365	gaaattaaaagcattgtacaaaatttaattgaacattccgaaaaagaaattatgaattgattgcg	2424
Oy	1741	aaagtaatggcaaaaggttaaatagatacaaaaattcaagaagcaggttaataataat	1800
Db	2425	aaattgattggcgaaggttaaatagatacaaaaattcaagaagcaggttaataataataat	2484
Oy	1801	caggaaatgaaataaagcattgagggttacccaataatatacaaaagcctattacattcaacgtg	1860
Db	2485	caggaaatgaaataaagcattgagggttacccaataatatacaaaagcctattatcaattcaacgtg	2544
Oy	1861	catataatagataatgccaaatatctgtaagaagtgtcttaataatctgaatgaatgaa	1920
Db	2545	catataatagataatgccaaatatctgtaagaagtgtcttaataatctgaatgaatgaa	2604
Oy	1921	aataaataatcaaaatgatactataaaaaaaggatcaacaattactttagtgatgaaatgaa	1980
Db	2605	aataaataatcaaaatgatactataaaaaaaggatcaacaattactttagtgatgaaatgaa	2664
Oy	1981	agattgcttcttaacgatactactcctcctaataatagctgtaacaaatalacacatcaagat	2040
Db	2665	agattgcttctttaaaccgattattactcctcctaataatagctgtaacaaatalacacatcaagat	2724
Oy	2041	gaagataatagaagaattcaataaaaggttatatggtccgaattcccggttctatatata	2100
Db	2725	gaagataatagaagaattcaataaaaggttatatggtccgaattcccggttctatatata	2784
Oy	2101	ctccatgagacctcaaaagggttgtagaattaaagaaatgataagtgaaaggtttalatacagaa	2160
Db	2785	ctccatgagacctcaaaagggttgtagaattaaagaaatgataagtgaaaggtttatatacagaa	2844
Oy	2161	tttgagaatgcgttgatagataatagctgtgataatctataatagaataagaaccaatcgattta	2220
Db	2845	tttgagaatgcgttgatagataatagctgtgataatctataatagaataagaaccaatcgattta	2904

QY	2221	gattcaaatctcaaaaaatcatctatctttaaaggaaggaaggaatctaacg	2280
Db	2905	ggttacaattcttcaaaaaattcattgattatttttaaggaagaaggaatttaacttcg	2964
QY	2281	tatggaagacaacaatgaagcgaattctttgcagaagcccttaagglttaagcatctaacg	2340
Db	2965	ttatggagagacaaatgcaacccgaattttttgcagaagcccttaagttaatgattctaacg	3024
QY	2341	gaccatgctgaacgtttaaaagltcaaaaaaatgctccgaaaacttccaattataac	2400
Db	3025	gaccatgctgaacgttttaaaagttcaaaaaaatgctccgaaaacttccaatttttaaac	3084
QY	2401	gattcagattaaagttcattatataaccctata	2430
Db	3085	gattcagattaaagttcattatataaccctata	3114
RESULT	5		
LOCUS	AF065404	181654 bp	DNA circular BCT 20-OCT-1999
DEFINITION	Bacillus anthracis virulence plasmid pXOI, complete sequence.		
ACCESSION	AF065404		
VERSION	AF065404.1	GI:4894216	
KEYWORDS			
SOURCE	Bacillus anthracis.		
ORGANISM	Bacillus anthracis		
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/staphylococcus group; Bacillus; Bacillus cereus group.		
AUTHORS	1 (bases 1 to 181654) Oknaka, R.T., Cloud, K., Hampton, O., Hoffmaster, A.R., Hill, K.K., Kelm, P., Koehler, T.M., Lamke, G., Kumano, S., Mahillon, J., Mantler, D., Martinez, Y., Rieke, D., Svensson, R., and Jackson, P.J. Sequence and organization of pXOI, the large Bacillus anthracis plasmid harboring the anthrax toxin genes J. Bacteriol. 181 (20), 6509-6515 (1999) 99445483		
TITLE	2 (bases 1 to 181654) Oknaka, R.T., Cloud, K., Hampton, O., Hoffmaster, A., Hill, K., Kelm, P., Koehler, T., Kumano, S., Lamke, G., Mantler, D., Martinez, Y., Rieke, D.O., Svensson, R. and Jackson, P. Direct Submission		
JOURNAL	Submitted (14-MAY-1998) Life Sciences Division, Los Alamos National Laboratory, TA43, LS-6, HRU-1, MS M888, Los Alamos, NM 87345, USA		
REFERENCE	Location/Qualifiers		
AUTHORS	1. 181654		
TITLE	/organism="Bacillus anthracis"		
JOURNAL	/strain="Sterne"		
REFERENCE	/db_xref="taxon:1392"		
AUTHORS	/plasmid="virulence plasmid pXOI"		
TITLE	complement(1127..1630)		
JOURNAL	/note="Hypothetical protein, Methanobacterium thermoautotrophicum, (AE000831)"		
FEATURES	/codon_start=1		
SOURCE	/transl_table=1		
CDS	/product="PXOI-01"		
	/protein_id="AAD32305.1"		
	/db_xref="GI:4894217"		
	/translation="MEVILFELILAVILPLINSVKKHVPKMKGKAGEKLYKRMLSKIDPKSYVILHNTVYETVEGDTQIDHIVAYAEVGAVVEKKNEGVIVGSEKARAYTGCHFRKSSRPNPHQNTKHKAIEMLIEQOLPCISMAAPHKCSLRVNVHSEKHKVKTLLHNKPAQ"		
	complement(1639..1644)		
	2751..2755		
	2760..3338		
RBS	/codon_start=1		
RBS	/transl_table=1		
RBS	/product="PXOI-02"		
RBS	/protein_id="AAD32306.1"		
RBS	/db_xref="GI:4894218"		
RBS	/translation="MLEHTILFSLVOLVPLIVSGIYILLETLDTEFMKTFMSEGGIFAIIFLIRFOYFQNKKSITISNQHKKMFPSKINIFLVSIFLVPASITPLKINATLDMFTVILFLLEDCLYDRDPFHNLYVKEIKOLKQKLTERTIIPNNEHGHTIKTRDQYFLTAVYLSGILPISSEVEEERNNLMGR"		

CDS
complement(3389..4024)
/codon_start=1
/transl_table=11
/product="px01-03"
/protein_id="AAD32307.1"
/db_xref="GI:4894219"
/translation="MKREKMGIVCLTVLMLILLVPLQAAQSVKVDENKLGNDVS
SFKRDLNNRTKTEKEDSDSGSSIIQIKIGIYAPARKGVNIFSIISFVGIYVMVAL
ITKNGOMKMGATGSMIFSTISMSMRLSYFLFSSVFSFPMNIFTFPLDFKATALLF
TTPMLTAGIRLRIRHESITQPEITRYMANNVITGSCILCMVLAVLPWLFANM"
complement(4024..4653)
/codon_start=1
/transl_table=11
/product="px01-04"
/protein_id="AAD32308.1"
/db_xref="GI:4894220"
/translation="MNINISNNVVEKLEQTVAREKKQVQLVSKSSNINAYMIFP
EFSEVKISMTLEDYEIENLSTSVKYVIDRKLSQLQEKONHSPIMTCEVEDVIDE
TIFHTYVKKPPTIANKVFKLVGCLGNRLDSLKEIHITLVNINNIRIEIEFVEMENET
VAVEKEIANNAMGAYTKANIYYAADLYAAEGNEIKLKSINKPLAGV"
complement(4034..4039)
complement(4660..4665)
complement(4989..5309)
/codon_start=1
/transl_table=11
/product="px01-05"
/protein_id="AAD32309.1"
/db_xref="GI:4894221"
/translation="MELLEYEDMYPOQEEVSIPTPEPEPVQOQVADVAKDMDFVMT
TVLQEOGLTRVNNNSNNKRYWIHTEGIELPENOLLVYMKLEVDSPSYTLVEMSNRO
LGN"
complement(5318..5323)
complement(5720..6220)
/codon_start=1
/transl_table=11
/product="px01-06"
/protein_id="AAD32310.1"
/db_xref="GI:4894222"
/translation="MIGDGLDLOTLILIONIKSGNOLKSGTEETKASGSKRYLANI
GNVITPVPESITQFTNMVYKLPETMDVAFKRAQVQKAVECTKREQVPGVAKONV
LEVDEIEDIIRALTTLTNDKRMVLVOKDITWDVSVIYVDSKRRKKQKNKIIDQY
EPDLFG"
complement(6231..6236)
complement(6544..8352)
/note="reverse transcriptase, IS629 (574 aa) E. coli
(AB011549): 306/543 positive aa (56%)
/codon_start=1
/transl_table=11
/product="px01-07"
/protein_id="AAD32311.1"
/db_xref="GI:4894223"
/translation="MNTVMKSAVSPQVTQMSIQMKEIENYVKLRQRIYAEQPN
KRRKRLQRLMRKANLILSRVYQINQGRKTAGIDITNTPEDRYKLHLKGY
SVRNTKAPVPRAYIPKKNKKRPLGIPVTKDRIFQNMVKNALSEQEGRFSPMSYGF
RRKSADHMANLPLKLSKGTNRAMIFEGDFQCFDNLNHEILSCIEGFPSMNIQ
MLNACICIDNKTFYKTEGTPQGGISPLANTALHGMKELGVRHFPFRDAMLYPD
SIGIYRVDADTVYVONSKEAEASMAKLOPYLDKGRGKLAKEKTRVYHITDFDLFGE
NPROYPTKEGSLFTKPSQSVKAKKESISELFSKSGISQILIRKLPLVTGILANY
WSVYAKVIYIGIDISYVHKVMHLKXKHHRGARINKKYHPDHTGVSDOKWLLTD
PDNHNOLIRMRWPIVIRVHLIKYRNSPDADMLKRYKARDEKIFNREFTNSKRILAK
KTYKCRVONSLVGEELPESNHIYKPVVIGKDEYDNLELLHSCSKOHALLLEWGN
GKQLPKVOAYLSOKIPIINSKKAVRMTLTFKKFKY"
complement(8362..8367)
complement(8956..11541)
/codon_start=1
/transl_table=11
/product="px01-08"
/protein_id="AAD32312.1"
/db_xref="GI:4894224"
/translation="MLQETIKODTDOQDOTPAIIQLEGTASHSAFCFAMATINHHREI
QPAVIGANDSTLKSFRRAISWGTSTFGEGRRESHSHTLDRKVVANVSGQEPFINTQ
TINKKAIASKEOLETKYIYVIAIDAEPMQVADLAPPGVGLPILEMAKPIYEEMLT
KNLQPLFVYEDKNEFSSISIAQVVLKERDCKEPLSEMIKSGCKQFPQGGTEKINKI

RBS
complement(11549..11554)
complement(11613..11840)
/codon_start=1
/transl_table=11
/product="px01-09"
/protein_id="AAD32313.1"
/db_xref="GI:4894225"
/translation="MSVLIGISGLIWMESNRMFRIATVLVICLASQLAGTHLE
VLEQSFVHGVDMLSAEOSLPEMKRFIDF"
complement(11847..11852)
complement(11856..12947)
/codon_start=1
/transl_table=11
/product="px01-10"
/protein_id="AAD32314.1"
/db_xref="GI:4894226"
/translation="MRLGNDLAAAGFPPTPLETGGKHLVRLLOMSEKRAYACFPCCGEG
TILRSPADFTKQAGVQFHTVGYELDANRYGOKAEOLDVYRSPSSPMIASHYFPLIF
LMPYVTELRTEKASKEKMEFNPFLKRAHLYLDEGGIMVYVITYDRARDISYTLAKN
YEEIGLMRGDENEEPEORFQCFVIGKRAITKNDYFNDRANFCEENSDIDFVAKH
VSTLQEMVNHKDMWTIPELRHQNKLIFTSVVDYKDAKYGIGKSEGSLAKRLNRNGY
LLOGETAERAKMPLIASQGLILITGVADLLEGDTLHAVRGSENVYEHSEBOL
EKTTEKVVLOKRRAKFITATPGVVELY"
complement(12956..12961)
complement(13209..13562)
/codon_start=1
/transl_table=11
/product="px01-11"
/protein_id="AAD32315.1"
/db_xref="GI:4894227"
/translation="MEDNOKSYAAPVMSSESRARVNRMGCEAEILIPDGKFOYFKP
HALNKEATQETALYIEGIOEKVGVIIIMRVGQGSFTSGRTIYAEPRKRLRLK
DYFDEEYVYEPKR"
complement(13574..13579)
14226..14231
14237..14929
/codon_start=1
/transl_table=11
/product="px01-12"
/protein_id="AAD32316.1"
/db_xref="GI:4894228"
/translation="MLKNLISISIVSESLIFVPIPOQDITMTKIVVIFNGLSISI
PITVARMQLOKFTLIDPKIKMKKEIKETNTVGLSLKLOKYOGLDKRSEYIN
RSEBHGILREKDKRKNATITIPINVAEKLIDYASLWLELDLHOKIYLVSY
VTFACEYLNALIPSLGSGSEKQEVAKLLSNIOKMWLVLSAGFEKVEKQDELYTE
FTFLVQOKNRG"

CDS
complement(15040..19002)
/note="erythrocyte invasion and possible binding protein;
Rhoprio protein (2401 aa), Plasmodium yoelii (U36927);
520/1194 positive aa (43%)
/codon_start=1
/transl_table=11
/product="px01-13"
/protein_id="AAD32317.1"
/db_xref="GI:4894229"
/translation="MAKKEIPDIKYEDIIRAKATGNLIIFVGAGISMLGLSPSMKG
YAEQLEDAIEKGVIDSLEYKLEDEPKKILITQMLFDEKEIIPKSANLEPFKEKN
EYKAIYERLYGMAIYITNTNDECLDVLTTPQSONEKYSMDLSEENKILEQSEET
KVFIDHTELLEAEKNGVNIHIGSGYKEESLVYVOYLVLENGTISQKSDHPSEVFL
DOVFNSTVVLVFMGGLFEFEELIYVLMSTKNSHTISIMYFESNENQSPQNTLYKK
YSNKGVELIPPNITKKNYDNLPIIDMSKVLGEVSGERDTSKTIISIDNMKCDDT

|||||
Db 129542 CTCGATGGACCTTCAAAAGTGTGAGAAATTAAGCAATGATAGTGGCTTTTATACACGAA 129601
QY 2161 ctctggacatgctgtgagatgattatgctgagatcatatagataagaaccaatctgattta 2220
Db 129602 TTTGGACATGCTGTGGATGATTTATCTGATATCTATTAGATTAAGAACCAATCTGATTTA 129661
QY 2221 gttacaattcttcaaaaatctatgattatctttaaagaagaaggagatctaactctg 2280
Db 129662 GTTACAAATTTCTAAAAAATTCATTGATATTTTAAAGAAAGGAGTAAATTTAACTTCG 129721
QY 2281 tatggagaacaacaatgaagcgaactcttttcgaagcccttagtttaatgattctacg 2340
Db 129722 TATGGGAGAACCAATGAACCGAATTTTTCGGAAGCCCTTTAGGTTAATGCAATTCACG 129781
QY 2341 gaccatgctgaacgctttaaaagttcaaaaaaaatgctcgaaaaacttccaaatttaaac 2400
Db 129782 GACCAATGCTGAACGTTTAAAGTTCAAAAAAATGCTCGAAACCTTCCAAATTTATTTAAC 129841
QY 2401 gatcagattaagtctatctatctaactataa 2430
Db 129842 GATCAGATTAACTTCATTATTAACCTATA 129871

RESULT 6
AX085496 1455 bp DNA PAT 09-MAR-2001
LOCUS DEFINITION Sequence 7 from Patent WO0112661.
ACCESSION AX085496
VERSION AX085496.1 GI:13275536
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequence.
1 (bases 1 to 1455)
REFERENCE Youle, R.J., Liu, X. and Collier, R.J.
AUTHORS Receptor-mediated uptake of an extracellular bcl-x l? fusion
TITLE protein inhibits apoptosis
JOURNAL Patent: WO 0112661-A 7 22-FEB-2001;
Department of Health and Human Services (US) ; PRESIDENT AND
FELLOWS OF HARVARD COLLEGE (US)
FEATURES
source Location/Qualifiers
1..1455
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="genetic fusion"
1..>1455
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAC34031.1"
/db_xref="GI:13275537"
/translation="MGSSHHHHHSSGLVPRGSHNAGHDVGMHVKKEKNKDNKR
KDERNKQOEHLKIMKHIVKIEVKGEEAVKAEKLEKVPDVEVMYKAGIGKI
YIVDDITKRIHSLKLEMDKKIKIDYKDLKHEHYVAKREPVYVIOSSDYVE
NTEKALNVYEIGKILSRDILSKINQPYKFLDYLNTIKNSDSDGDLFLFNOLKEH
PTDSVEFLKONSNEVQVFAKFAVYIEPOHVDIOLYAPAEVYMKFNOETINS
MSQNRRLVDFPLSYKLSOKGYSWQSFEDVENRTEABDGESEMETSAINGNSMH
LADSPAVNGATRAHSSLDAREYITPAAYKQAKREKDELELTKRAFSDLTSQLHTP
GTATQSFQDVVNELEKRDGVNMGRIYAFSPGALCIVESVDKEMOVLVSRIAMMATYL
NDHLEPMIQEONGMDTFVELYGNNAAESRKGOER"

BASE COUNT 487 a 266 c 369 g 333 t
ORIGIN

Query Match 31.6%; Score 767.6; DB 6; Length 1455;
Best Local Similarity 99.5%; Pred. No. 7.7e-100;
Matches 770; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 94 cagggggcggggcggtcattgattagtgatgcacgttaagaagaagaataa 153
Db 58 CATATGGCGGGGTCATGCTGATGATGACGTAAGAGAGAGAAAAATATA 117

QY 154 gatgagaataagaagaagaatgagaacggaataaacaacaggagaatcttaagga 213
Db 118 GATGAGAAATTAAGAGAAAGATGAGAACGAAATTAACACAGGAAGCATTTTAAAGGAA 177
QY 214 atcatgaacacatctgtaaaatagaagtaagaaggaggaagctgttaaaaaagagca 273
Db 178 ATCATGAACACATTTTAAATAATAGAGTAAGGAGGAGAGACCTGTTAAAGAGGCA 237
QY 274 gcaagaagctactctgagaagatcacatctgactgttctttagagatgataaagcaatgga 333
Db 238 GCAGAAAGCTACTTATGAGAAAGTACATCTGATCTTTAGAGATGATTAAGCAATTTGGA 297
QY 334 ggaagaatataatctgtgagatgagatatacaacaacatatactttagaagaatcatc 393
Db 298 GGAAAGATATATATGCTGATGATGATTAACAAACATATATCTTTAGAACGATTAATCT 357
QY 394 gaagataagaagaagaataagaacatcttctgggaagaatgcttataatgaacatct 453
Db 358 GAGGATTAAGAAATAATTAAGACATTTATGGAAAGATGCTTTATTAACATGAACATTTAT 417
QY 454 gtaatacaaaagaagaatatacgaaaccgtactgttaatacactcgcgaagaatata 513
Db 418 GATATGCAAAAGAGAGATATGAAACCCGTAATCTGTAATCCAATCTTCGAAAGATTAATGTA 477
QY 514 gaaataactgaaagaagcactggaacgttattatagaatagtgtaagatatacaaggaat 573
Db 478 GAAATATCTGAAAAAGGCACTGAACGTTTATTAAGAAATAGTAAGATATTAATCAAGGAT 537
QY 574 atttaagaataaataatcaacacatacgagaatctttagatgatacaacacataa 633
Db 538 ATTTTAAGTAAATTAATCAACCATATCAAGAAATTTTATGATGATTAATACCATTAAN 597
QY 634 aatgatactgatacagatgagaagaatcttataatctaatacagcttaagaagaatccc 693
Db 598 AATGCAATCTGATTCAGATGACAGACATCTTTTATTAACCAACGCTTAAGAACATCC 657
QY 694 acagaaactctctgtagaactcttggagaacaaatagaagaatgaggtagaagaagatttgg 753
Db 658 ACAGACTTTCTGTAGAAATCTTGGACAAATAAGCAATGAGGTACAAAGCATTTTGGC 717
QY 754 aaagctctgcatatataatcgagccacagatcgatgattttaaagctttagacgg 813
Db 718 AAAGCTTTTGCATATTAATTCAGGACACAGCATCGATGATTTTACAGCTTTATGACCG 777
QY 814 gaagctttaaatacagatgataaatttaacgaagaagaataatcatccttg 867
Db 778 GAAGCTTTTAATTAACATGATTAATTAAGCAACAGAAATTAATCATTCATG 831

RESULT 7
I33397 1368 bp DNA PAT 06-FEB-1997
LOCUS DEFINITION Sequence 5 from patent US 5591631.
ACCESSION I33397
VERSION I33397.1 GI:1824188
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1368)
AUTHORS Leppla, S.H., Klimpel, K.R., Arora, N., Singh, Y. and Nicholls, P.J.
TITLE Anthrax toxin fusion proteins, nucleic acid encoding same
JOURNAL Patent: US 5591631-A 5 07-JAN-1997;
FEATURES
source Location/Qualifiers
1..1368
/organism="unknown"
ORIGIN

BASE COUNT 411 a 316 c 358 g 283 t
ORIGIN

Query Match 31.4%; Score 762.4; DB 6; Length 1368;
Best Local Similarity 99.2%; Pred. No. 4.2e-99;
Matches 766; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy	100	gcgggcccgttcacatggtgatcgtatgcacgttaaaagaaggaagaaaataaagtacg	159
Dd	1	GGCGGCGCTCATGGTGAATGTAGTATGCATAAAGAAGAAAGAAAATTAAGATGAC	60
Oy	160	aataagagaaaaagatctgaagacgcaaatlaaacacagggaagacattlaaaggaaatcatg	219
Oy	220	aaacacattgtaaaaaatagaagtlaaadaggagggaagccgttlaaaaaaagcgacgaa	279
Dd	121	AAACCACTTTGTAAAAATGTGAAGTAAAAAGGGAGGAGAACCTGTTAAAAAAGAGCAGCAA	180
Oy	280	aagcactctggaaagaatcacctccatgattgtttgaagtgatcataagcaattggaagaag	339
Dd	181	AAGCTACTTGAGAAAGTAGCCATCTGATGTTTTAGAGATGTATAAAGCAATTGGAGGAAAG	240
Oy	340	alatalatttgtgga tbgatcattacaacacatatccttagaagcattatcctgaagat	399
Dd	241	ATATATATTGTGGATGGTGATATTACAACATATATCTTTAGAGCATTTATCTGAAGAT	300
Oy	400	aagaaaaaaataaagacatttatgggaagaatgcttattacatgacatttlylat	459
Dd	301	AAGAAAAAAAAAAGACATTTATGGGAAGATCCTTTATTACATGAACATTTATGTATAT	360
Oy	460	gcaaaagaagaatctgaaaccgcacctgtgtatccaaccttcogaaagttitgtagaanaat	519
Dd	361	GCAAAAAGAGATATGAACCCGTCCTTGTAATCCAATCTTCGGAAATATATGTAGAAAT	420
Oy	520	actgaaaaagcagcctgaacqgttattatgaaaatgaagtaagatatatlatacgaagatat	579
Dd	421	ACTGAAAAGCACAGTGAACGTTATTATGAATAGGTAAAGATATATCAAGGATATTTTA	480
Oy	580	agtaaaatlaatcaacacatatcagaanaattlttagatgtalttaaataccattaaaatgca	639
Dd	481	AGTAAAAATTAATCAACCATTATCAAAAAATTTTNAATGTATTTAAATACCATTAATAATGCA	540
Oy	640	tctgtattcaga tggacaagatccttttattacaaacagctttaagaaatccacagac	699
Dd	541	TCTGATTACAGTAGGCAACAGATCTTTTATTACTATATAGCTTANAGAACATCCACAGAC	600
Oy	700	tctctgtagaatctcttggaaacaaatlagcaatlgagtlacaaagaagtatttcgaaagc	759
Dd	601	TTTTCTGTAGAAATCTTGGAACAATAATAGCAATGAGGTACACAGATTTGGGAAAGCT	660
Oy	760	tttgcattatattatccgagccagcgatcgtgtagtlttacagctttagcaccggaagct	819
Dd	661	TTTGCAATATTATATCGAGCCACACCAATCGGATGTTTTACAGCTTTATGACCCGGAAGCT	720
Oy	820	tttaattacatggaataaattlaagcaacaaagaataaataatcatccttggaag	871
Dd	721	TTTATATTCATGATGAATTAATTTAACGAACAAGAAATATATCTACTCGGCGACG	772
RESULT	8		
LOCUS	169375	1368 bp DNA	
DEFINITION	Sequence 5 from patent US 5677274.	PAT	04-FEB-1998
ACCESSION	169375		
VERSION	169375.1	GI:2831497	
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 1368)		
AUTHORS	Leppla,S.H., Klimpel,K.R., Arora,N., Singh,Y. and Nichols,P.J.		
TITLE	Anthrax toxin fusion proteins and related methods		
JOURNAL	Patent: US 5677274-A 5 14-OCT-1997;		
FEATURES	Location/Qualifiers		
SOURCE	1..1368	/organism="unknown"	
BASE COUNT	411 a 316 c 358 g 283 t		
ORIGIN			

Query Match	Similarity	31.4%	Score 762.4	DB 6	Length 1368
Best Local	Similarity	99.2%	Pred. No. 4.2e-99		
Matches	766	Conservative	0	Mismatches	6
				Indels	0
				Gaps	0
Qy	100	gcggcgagtcacatgcatgtagtgaatgaacgtgaagaagaaagaaagaaataaagatgag	159		
Db	1	gcggcgagtcacatgcatgtagtgaatgaacgtgaagaagaaagaaataaagatgag	60		
Qy	160	ataagagaaaaagatgaaagaaacgaataaacaacggaagacatttaaggaaatcatg	219		
Db	61	ataagagaaaaagatgaaagaaacgaataaacaacggaagacatttaaggaaatcatg	120		
Qy	220	aaacacattgtaaaaataagaagtaaagggggaagacgtgttaaaaaaaggacgaagaa	279		
Db	121	aaacacattgtaaaaataagaagtaaagggggaagacgtgttaaaaaaaggacgaagaa	180		
Qy	280	aagctacttgagaaagatcacatctgactgttctagaagatgaataaagcaattgagaaag	339		
Db	181	aagctacttgagaaagatcacatctgactgttctagaagatgaataaagcaattgagaaag	240		
Qy	340	atactatttggatggatggatatactcaaacatatactcttgaagacattatctgaagat	399		
Db	241	atactatttggatggatggatatactcaaacatatactcttgaagacattatctgaagat	300		
Qy	400	aagaaaaaataaagaacatttatggaagaagatgcttatactatgacaatcatatgatat	459		
Db	301	aagaaaaaataaagaacatttatggaagaagatgcttatactatgacaatcatatgatat	360		
Qy	460	gcaaaagaagaatatagaacccgactctgttaatccaaactcttoggaaatatagtagaat	519		
Db	361	gcaaaagaagaatatagaacccgactctgttaatccaaactcttoggaaatatagtagaat	420		
Qy	520	actgaaagagcactgaacgttattatgtaaataggtaagatactcaagggatatttca	579		
Db	421	actgaaagagcactgaacgttattatgtaaataggtaagatactcaagggatatttca	480		
Qy	580	agtaaaatlaatacaacatatacgaatctttagatgataataacatlaaaaaatgca	639		
Db	481	agtaaaatlaatacaacatatacgaatctttagatgataataacatlaaaaaatgca	540		
Qy	640	ttctgaatgaatgagcaagaatcttatttactaatacagcttaagaacatccacaagac	699		
Db	541	ttctgaatgaatgagcaagaatcttatttactaatacagcttaagaacatccacaagac	600		
Qy	700	ttctctgtagaatctcttggaacaaatagcaatgagatgacaaagatcttctgcaagaact	759		
Db	601	ttctctgtagaatctcttggaacaaatagcaatgagatgacaaagatgacaaagact	660		
Qy	760	tttgcatattatatacagacacagacatctgtagatgttctaacgcttatatgacacggaagct	819		
Db	661	tttgcatattatatacagacacagacatctgtagatgttctaacgcttatatgacacggaagct	720		
Qy	820	tttaattcaatgagataaattcaacgaacgaagaataaatactatcccttgaag	871		
Db	721	tttaattcaatgagataaattcaacgaacgaagaataaatactatcccttgaag	772		
RESULT	9				
LOCUS	133398	1425 bp	DNA	PAT	06-FEB-1997
DEFINITION	Sequence 7 from patent US 5591631.				
ACCESSION	133398				
VERSION	133398.1	GI:1824189			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 1425)				
AUTHORS	Leppia,S.H., Klimpel,K.R., Aroa,N., Singh,Y. and Nicholls,P.J.				
TITLE	Anthrax toxin fusion proteins, nucleic acid encoding same				
JOURNAL	Patent: US 5591631-A 7 07-JAN-1997;				

ACCESSION	133399				
VERSION	133399.1	GI:1824190			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
1 (bases 1 to 1524)					
AUTHORS	Leppia,S.H., Klimek,K.R., Arora,N., Singh,Y. and Nicholls,P.J				
TITLE	Anthrax toxin fusion proteins, nucleic acid encoding same				
JOURNAL	Patent: US 5591631-A 9 07-JAN-1997;				
FEATURES	Location/Qualifiers				
source	1..1524				
BASE COUNT	436 a	374 c	415 g	299 t	
ORIGIN					

Query Match	31.4%	Score 762.4	DB 6	Length 1524
Best Local Similarity	99.9%	Pred. No. 4,2e-99		
Matches 763; Conservative	0	Mismatches 1	Indels 0	Gaps 0

QY	100	gcggcgccgtccatgcatgcatgtagtctgcacgtcaagaagaaagaaataaagatgag	159
Db	1	CGCGCGCGGTCAATGCGATGCTAGGTTCGACGCTAAAAGCAAAAGAGAAAATTAAGTTCAG	60
QY	160	aataagagaaaaagatlgaaagaaacgaaaaataaacacaggaagagcatltaaggaaatcatg	219
Db	61	AATAAGACAAAAAGATGAAAGAACGAATATAAACACAGGAAGACCATTTAAAGGAATTCATG	120
QY	220	aaacacatltgtaaaaaatagaagtaaaagggaggaagcgtgttaaaaaaaggcagcagaa	279
Db	121	AAACCATTTGTAATAAATGGAAGTAAAAGGCGAGGAAGCTGTTAAAAAAGAGCGACAGCAAA	180
QY	280	aagcgaacttggaaagtaaccatctgcatgtlcttagagatggtataaagcaattggaggaaag	339
Db	181	AAGCTACTTGGAGAAAGTCCATCTCATGTGTTTAAAGATGTATTAAGCAATTTGGAGGAAG	240
QY	340	atatacatlttgtagtgcgtgcatatlaacaaacataatctlttagaagcattactctgaagat	399
Db	241	ATATATATTGTGGATGGTGATATTACAAACATATATCTTTGAAGCAATTAICTGAAGAT	300
QY	400	aagaaaaaaataaaagacattatgvggaagaatgcttattatcatgaaacattatglatat	459
Db	301	AAGAAAAAAATPAAAAAGACATTTATGTGGAAAGATGCTTTATTCATGAAACATTATGTATAT	360
QY	460	gcaaaagaaggtatgaaacccgtacttctttaaaccacatcttggaaagattatlagaaaat	519
Db	361	GCAAAAGAAAGATATGAACCCGCTACTGTGTAATCCAAATCTTCGGAAGATTATGTAGAAAT	420
QY	520	actgaaagaagcactgaaacgttattcatatgaaataggtaagatactlaaaggatattta	579
Db	421	ACTGAAAAGGCACGTACAGCTTTATATGGAATAGGTAAATTTTTCAAAGGATATTTTA	480
QY	580	agtaaaatlaatacaaccatctcagaagaattttagatlglatlaataacca tlaaaaatgca	639
Db	481	AGTAAAAATTAATCAACCATATCAGAAATTTTATAGATGTATTAATAATCACTATPAAAAATGCA	540
QY	640	ctctgatcgaatgagcaagaatcttlaatttactaatcagcttaaggaaataccacacagac	699
Db	541	TCTGATTTAGATGCGCAAGATCTTTTATTTACTTAATCAGCTTTAAGGAACATCCCAACAGAC	600
QY	700	ttctctgtagaattcttggaaacaataatagcaatgtagtcaagaagaatttgcgaagct	759
Db	601	TTTTCTGTAGAAATCTTGTGGAACAAAATATAGCAATGAGTTCANAGAAATATTTTCGAAAGCT	660
QY	760	tttgcataatlatatcgaccacacagatcgtgtgatttcaagcttlatgcacggagaagct	819
Db	661	TTTGCATATTTATATGACGCCACACAGCATCGTGATGTTTACACTTATATGCACCGGAAGCT	720
QY	820	tttaatacatcaggtataaaattcaagcaacaagaataaatactctc 863	
Db	721	TTTAAATTCATGATGATAAATTTTAACAACAAGAAAATAAATCTTATAC 764	

RESULT	12				
LOCUS	169377				
DEFINITION	Sequence 9 from patent US 5677274.				
ACCESSION	169377				
VERSION	169377.1				
KEYWORDS	GI:2831499				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 1524)				
TITLE	Lepple, S. H., Kimpel, K. R., Atora, N., Singh, Y. and Nichols, P. J				
JOURNAL	Antirax toxin fusion proteins and related methods				
FEATURES	Patent: US 5677274-A 9 14-OCR-1997;				
	Location/Qualifiers				
source	1..1524				
	/organism="unknown"				
BASE COUNT	436 a	374 c	415 g	299 t	
ORIGIN					

Query Match	31.4%	Score 762.4;	DB 6;	Length 1524;
Best Local Similarity	99.98%	Pred. No. 4.2e-99;		
Matches 763; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

Oy	100	gogggcggtctatggtatgtagttagtgaagtaaaagaaaggaagaaataaagatgag	159
Db	1	gogggcggtctatggtatgtagttagtgaagtaaaagaaaggaagaaataaagatgag	60
Oy	160	ataagagaaaaagatlgaaacgaaataaaacacaggaagagcatltaaggaatacatg	219
Db	61	ATAAGAGAAAAAGATGGAAGACGAAATPAAACACAGGAGAGCAATTTAAAGGAATCATG	120
Oy	220	aaacacatctgtataaaatagaagttaaaggggaaggaagctgttaaataaagggcagcgaa	279
Db	121	AAACCATTTGTAAAAATTGAAGGTAAAGGGAGAGAGCTGTAAAAAAGAGCGACGAA	180
Oy	280	aagctacttggaaagtaaccactctgtagtctttagagatgltataaagcaatttggaggaag	339
Db	181	AAGCTACTTGGAAAAAGTGCACCTCGATGCTTTTAAAGATGTATTAAGCAATTGGAGAAAG	240
Oy	340	ataatatcttgggtatggtgatataataaaacatatctttagaagcatatctcgaagc	399
Db	241	ATATATATTGTGGATGGATGGATTTACAAACATATATCTTTAGAAAGCATTTATCGAAGAT	300
Oy	400	aagaaaaataaaagacatttcatgggaagaatgcttatcatcagacatlatglat	459
Db	301	AAGAAAAAATTAAGACATTTATGGGAAGAGCTTTATTACATGAAACATTATAGTATAT	360
Oy	460	gcaaaagaagatatgaaccggtactctgttaatccaatcttcggaagatlatgtagaat	519
Db	361	GCAAAAGAGGATATGAAGCCCGTACTCTGTATATCCAAATCTTCGGAAGATTATGTGAAAT	420
Oy	520	actggaagagcagcgaacgtttatataagaataggtaagatataatcaaggaatattia	579
Db	421	ACTGAAAGGACGACGTCGACTTTATTATGCAATTAGTAAAGATATTATCAAGGATATTTTA	480
Oy	580	agtaaaatlaatcaaccalacgaagaattttagatglatbaaataccatltaaaatgca	639
Db	481	AGTAAAAATTAATCAACCATATACAAAAATTTTAAATGATATTAATTAACCATTTAAATGCA	540
Oy	640	tctatcttaagttagcaaatcttttatcttaataagcttaaggaacatccaagagc	699
Db	541	TCTGATTAAGATGGCAACGATCTTTTATTACTATACAGCTTTAAAGAACATCCACAGAGC	600
Oy	700	ttttctcagaatctcttggaaacaaatagcgaatgagtlacaagaagatcttgcgaagct	759
Db	601	TTTTCTGTAGAAATCTTGGAACAAAAATAGCAATGAGGACAAAGATATTTGCCAAAGCT	660
Oy	760	tttgatatattatcgcagccacagacatcgtgtagtctttaaagctttatgcacgggaagc	819
Db	661	TTTGATATTTATATCGACCCACACATCGATATGTTTTTACAGCTTTATGACCGGAAGCT	720

RNHKTEKNTEKEREKFKDSINNVLKTEFTNETLTKIOOTODLLKPKIVLEIYSELG
EIIYFTDIDVEHKELODISEEKNMSRGEKPFARFEKREPTKLINIKDVA
INSEOSKEVEYIEHKEGLSDISDKSLDEPDLNLSLSDSSDLSLFSOKREKL
ELNNKSIDINFIKENLTERPOHAFSLASYFPADHRVLEIYAPDMFEYMKLEKGF
EKISSELKREGEVDRIIDLGEKALASGLVPEHDAFKIAELWYILFRVNL
ATNLIKSGVATKGLNVDKSSDMGPVAGYIPFDODLSKKHQOOLAVKGNLENKKSIT
EHGEIETKIPKLIDHLRIEELKENGIIILKKEIDNKKYLLLESNOVYEFRI SDN
NEVOYKTEKGIIVLGEKFMNRNIEWAKNVEGVILKPLTADYDLFALAPSTETIKKOI
POKEMDKVYVNPNSLEKOKGLTNILITVIGIERKPDSTKGLTSMNOKMDRLNNAVY
TYTGTDGVNHTGEODNEPEPEKDNIEFTINPESGFILTKNMENTGFTIEKNTGKYD
LYTFNRSYNKIAPGNKAYIEWTDPIITAKINTIPTSAEFLIKNLSISRRSSNWGYKXS
GDDEFAKESKVIKAGYLSDYNSANHIFSOEKKRISIFRGIOAVNEIENVLKSKO
IAPYKNVFOYLKERITNOVOLLTHOKSNIEFKLYKOLNFTENETDNFEVFOKIID
EK"

BASE COUNT 1253 a 331 c 516 g 890 t
ORIGIN

Query Match 6.7%; Score 162.6; DB 6; Length 2990;
Best Local Similarity 52.7%; Pred. No. 3e-14;
Matches 404; Conservative 0; Mismatches 354; Indels 9; Gaps 2;

OY 131 taaagagaagaaagaaataagatgagaaataagagaagaaagaaagaaataaa 190
D 523 TGAATGAACATTACACGAGAGATGATATTAAGAAACCATTAACCTGAAAAAATAAA 582
OY 191 cagagagaagcatttaagaaatacctgaacacattgtaaaatagaaataaagagg 250
D 583 CTGAAAAAGAAAAATTTAAAGACAGATTTAACTTAAGTTAAACAGAAATTTACCAATG 642
OY 251 aggaagcgtttaaaaaagagcagagaagaactacttgagaagtaaccatctgatt 310
D 643 AAATTTAGATTAATTAACGACAGACACAGACTTATTAAGTAAGATGATGAC 702
OY 311 tagagatgataaagcaatttgaggaagaatataattggtgagtgatattacaac 370
D 703 TTGAAATTTAGTGAATTAAGAGAGAAATCTATTTACAGATATGATTTAGTAGAAC 762
OY 371 atatacttgagaagcattatcgaagaataaagaataaagaataacattatggaag 430
D 763 ATTAAGAGATTACAAAGATTTAACTGAAGAGAAAAATGATGATGATGAGGTGAAA 822
OY 431 atgcttattacatgaacaattatgatalgcaaaagaagaatataaccgtactgt 490
D 823 AAGTTCGTTTGCATCCCGTTTGTATTTGAAAAAGAAAGGAAACACCTAAA---TTAA 879
OY 491 tccaactctcggaagattatgtagaataactgaaagcactgaagcttattatga 550
D 880 TTATTAATATCAAGATTTATCAATTTAATGTAACAAAGTAAAGAAAGTATATGAAA 939
OY 551 tagttagaatacttataaggaattttaagtaataatcaacatatacagaatttt 610
D 940 TTGGAATGAGGATTTCTCTTATATTAATTAAGTAAGGATTAATCTTAAATCCAGATTT 999
OY 611 tagatgattataacatcaataaataaataacatgattcagatgagaagaacttt----- 665
D 1000 TAAATTTAATTAAGAGTTTAAGCGATGATGATGATGATGATGATGATGATGATGATGAT 1059
OY 666 -attactaatcagccttaagaaacatccacagacattctgttagaactcttgagaacaa 724
D 1060 AATTTAAAGAACACTGAATTAATTAAGTAATTAAGTAATTAAGTAATTAATTAAGTAATTA 1119
OY 725 atagcaatgagtgatacaagaaglatitggaagaagcttgcattatatttcgacacagc 784
D 1120 ATTAACTGAATTTACAGATCGCTTTCTTTAGCGTTTCTTATATTTTGACACTGAC 1179
OY 785 atcgtgatttcaacacattacacacgaagccttcaatatacattgataaataa 844
D 1180 ATTAAGAGGATTTAGAGATTATATGCCCCGACATGTTTGAGATATTAAGTAAGTTAGAAA 1239
OY 845 aacaagaataaatacatalccttgagaagaacttaagaatcaacgagtg 891
D 1240 AAGGCGATTTGAGAAATTAAGTGAAGTTGAGAAAGAAAGAGGTG 1286

RESULT 15
LOCUS BACCYA 3420 bp DNA BCT 12-OCT-1995
DEFINITION Bacillus anthracis edema factor gene (cya) gene, complete cds.
ACCESSION M24074
VERSION M24074.1 GI:142812
KEYWORDS adenylate cyclase; anthrax toxin; cya gene; edema factor.
SOURCE Bacillus anthracis (individual isolate Sterne) DNA.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/staphylococcus group; Bacillus/Bacillus cereus group.
REFERENCE 1 (bases 1 to 3420)
Robertson,D.L., Tzipetz,M.T. and Leppla,S.H.
Nucleotide sequence of the Bacillus anthracis edema factor gene (cya): a calmodulin-dependent adenylate cyclase
Gene 73 (2), 363-371 (1988)
COMMENT The authors depositing this sequence (M24074) consider that it contains several errors. M23179 is considered to be more correct.
FEATURES
Location/Qualifiers
1..3420
/organism="Bacillus anthracis"
/isolate="Sterne"
/db_xref="taxon:1392"
544..642
/gene="cya"
544..2946
/gene="cya"
/transl_start=1
/transl_table=1
/product="edema factor"
/protein_id="AA09213.1"
/db_xref="GI:142813"

BASE COUNT 1400 a 393 c 593 g 1034 t
ORIGIN

Query Match 6.7%; Score 162.6; DB 1; Length 3420;
Best Local Similarity 52.7%; Pred. No. 2.9e-14;
Matches 404; Conservative 0; Mismatches 354; Indels 9; Gaps 2;

OY 131 taaagagaagaaagaaataagatgagaaataagagaagaaagaaagaaataaa 190
D 644 TGAATGAACATTACACGAGATGATATTAAGAAAGAACCATTAACCTGAAAAAATAAA 703
OY 191 cagagagaagcatttaagaaatacagaaacacattgtaaaatagaaataaagagg 250
D 704 CTGAAAAAGAAAAATTTAAAGACAGTATTAATTAAGTTAAACAGAAATTTACCAATG 763
OY 251 aggaagcgtttaaaaaagcagagaagaactacttgagaagataccatctgatttt 310

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:48:20 : Search time 112.1 Seconds

(without alignments)
181.053 Million cell updates/sec

Title: US-09-747-521-2_COPY_9_282

Perfect score: 1404

Sequence: 1 KVISMSCLVATILSGPVFI.....RDVLQVLAPEAFNVMKENE 274

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: A_Geneseq_1101.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1404	100.0	809	22	AAAB7305
2	1282	91.3	456	15	AAAB60180
3	1282	91.3	472	15	AAAB60181
4	1282	91.3	485	22	AAU00222
5	1282	91.3	508	15	AAAB60182
6	1282	91.3	776	15	AAAB60178
7	427	30.4	800	11	AAAR04236
8	129.5	9.2	980	21	AAAB18294
9	129	9.2	1245	21	AAAB18244
10	125	8.9	558	21	AAAB18273
11	122.5	8.7	508	21	AAAB18187

12	122	8.7	481	20	AAV20047	B. burgdorferi ant
13	122	8.7	497	12	AAV20046	B. burgdorferi ant
14	119	8.5	1427	12	AAAB10534	Human 160kD mediat
15	118.5	8.4	652	18	AAAB18010	Plasmodium falcipa
16	118	8.4	687	19	AAAB1586	Truncated restin p
17	118	8.4	1392	20	AAV06999	Restin protein seq
18	117	8.3	1351	21	AAAB18290	Plasmodium falcipa
19	115.5	8.2	561	19	AAAB3043	Streptococcus uber
20	114	8.1	1135	21	AAV84460	Amino acid sequenc
21	113	8.0	477	20	AAV34601	Chlamydia pneumoni
22	113	8.0	593	14	AAAB4669	Ap SeroType 7 65kD
23	113	8.0	593	21	AAV51774	A. pleuropneumonia
24	113	8.0	593	21	AAV80378	Drosophila kinesin
25	113	8.0	975	19	AAW72746	Plasmodium falcipa
26	113	8.0	1979	21	AAAB18171	Plasmodium falcipa
27	113	8.0	2440	18	AAW20828	H. pylori cytoplas
28	112.5	8.0	1192	21	AAAB18165	Plasmodium falcipa
29	112	8.0	1093	19	AAW40540	Mutant C-beta prot
30	112	8.0	1558	21	AAAB18324	Plasmodium falcipa
31	112	8.0	1643	21	AAAB45564	Virulence gene pro
32	112	8.0	1786	18	AAW24790	P. falciparum live
33	112	8.0	2110	21	AAAB45528	Virulence gene pro
34	111	7.9	700	16	AAAB75731	B. burgdorferi str
35	111	7.9	716	14	AAAB30730	B. burgdorferi 79
36	111	7.9	975	22	AAAB94042	Human protein sequ
37	111	7.9	1164	17	AAAB5781	Group B streptococ
38	111	7.9	1164	19	AAW40537	Group B streptococ
39	111	7.9	1164	21	AAV84459	Amino acid sequenc
40	110.5	7.9	228	20	AAV20077	B. burgdorferi ant
41	110.5	7.9	259	20	AAV20076	B. burgdorferi ant
42	110	7.8	1164	19	AAW40541	Mutant C-beta prot
43	110	7.8	2485	21	AAAB18172	Plasmodium falcipa
44	109	7.8	265	22	AAAB1187	Human polypeptide
45	109	7.8	307	22	AAAB1714	S. epidermidis ope

ALIGNMENTS

RESULT 1
ID AAB47305 standard; Protein: 809 AA.
XX AAB47305;
XX 29-AUG-2001 (first entry)
DE Wild type B. anthracis lethal factor.
XX Lethal factor; LF: immunogen; LF4: protective antigen; PA: DNA vaccine.
KW humoral; cell-mediated; immune memory response.
XX Bacillus anthracis.
OS
XX Key Location/Qualifiers
FH Peptide 1..33
FT /label= Signal peptide
FT /note= "Not given in the specification"
FT Protein 34..809
FT /label= LF
FT Peptide 42..285
FT /label= LF4
XX WO200145639-A2.
PN 28-JUN-2001.
PD LFN-BCL-XL Apoptos
XX LE(1-254)--TR--PE
PF Lethal factor of B
XX Adenyl cyclase gen
XX Plasmodium falcipa
PR Plasmodium falcipa
XX 22-DEC-1999; 99US-0171459.
PA (OHIS) UNITV OHIO STATE RES FOUND.
PA (GALL/) GALLOWAY D R.

KM Pseudomonas; exotoxin.
XX
XX Bacillus anthracis.
OS
OS Pseudomonas sp.
XX
XX WO9418332-A.
XX
XX 18-AUG-1994.
XX
XX 14-FEB-1994; 94WO-US01624.
XX
XX 12-FEB-1993; 93US-0021601.
PR 25-JUN-1993; 93US-0082849.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Arora N, Kilmpel K, Leppla SH, Nichols PJ, Singh Y;
XX WPI: 1994-279753/34.
DR N-PSDB: AAQ70182.
XX
XX Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumour cells
PT or HIV-infected cells
XX
XX Example 2; Page 90-92; 124pp; English.
XX
XX This sequence is a fusion protein comprising amino acid residues
CC 1-254 of the anthrax protective antigen binding domain of the native
CC anthrax lethal factor, a two residue linker and residues 398-613 of
CC a Pseudomonas exotoxin A activity inducing domain of a second
CC protein. Such fusion proteins may be useful for the specific
CC killing of tumour cells or the killing of cells infected with
CC intracellular pathogens, especially HIV, depending on their
CC components.
XX
XX Sequence 472 AA;
SQ

Query Match 91.3%; Score 1282; DB 15; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.6e-93;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGGHDVGAMHVKEKRNKDEENKRDDEENKTOEHLKEIMKHIVKTEVGEAVKKEAAE 85
DB 4 agghdvgamhvkekeknkdenkrkdeernktgeehkeltmkhivklevgeevkkeaee 63
QY 86 KLEKVPDVLBYKRAIGKRIYVDDITKHTLSLEALSDKKRKIKDIYGRDALHEHYV 145
DB 64 kilekvpdvlbykraigkriyvdditkhtlslealسدkkkkikdiygdalldhehyv 123
QY 146 AKEGIEPVLYVITOSSEDYVENTERKALNVYIEGKILSRDILSKINOPYQKFLDVLNTIKNA 205
DB 124 akegiepvlvitosseedyventekalnvyielgkilsrdilskinpqyqkfldvlnltikna 183
QY 206 SPSSDODLFTNQLKEHPTDFSEVFLEQNSNEMOEVEFAKAFYUITEPORDVLTQCAPPA 265
DB 184 sdsdsdgdllftnqlkehptdfsevfleqnsnevgevfakafayltepordvltqlyapapa 243
QY 266 FNYMDKENE 274
DB 244 fnymdkene 252
XX

RESULT 4
AAU00222
ID AAU00222 standard; Protein; 485 AA.
XX
XX AAU00222;
AC
XX
XX 31-MAY-2001 (first entry)
XX Lfn-Bcl-XL apoptosis-modifying fusion protein.

XX Human; Lfn-Bcl-XL; apoptosis; cancer; spinal muscular atrophy;
KM anthrax lethal factor; neoplasm; tumour; hyper-proliferation;
KM Alzheimer's disease; neurodegenerative disorder; stroke;
KM transient ischaemic neuronal injury; spinal cord injury;
XX Huntington's disease.
XX
XX Chimeric - Homo sapiens.
OS Chimeric - Corynebacterium diptheriae.
OS Chimeric - Synthetic.
XX
XX Key Location/Qualifiers
FH Region 5..10
FT /note="6x histidine tag"
FT Region 21..276
FT /note="Anthrax lethal factor amino acids 1 to 255"
FT Region 277..485
FT /note="Bcl-XL amino acids 1 to 209"
XX
XX WO200112661-A2.
XX
XX 22-FEB-2001.
XX
XX 15-AUG-2000; 2000WO-US22293.
XX
XX 16-AUG-1999; 99US-0149220.
XX
XX (HARD) HARVARD COLLEGE.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Youle RJ, Liu X, Collier RJ;
DR WPI: 2001-218343/22.
DR N-PSDB: AAS00250.
XX
XX Novel fusion protein for modifying apoptosis in target cell and
PT reducing apoptosis after transient ischaemic neuronal injury, has two
PT domains which targets protein to a cell and modifies apoptotic response
PT of cell
XX
XX Claim 4; Page 64-65; 65pp; English.
XX
XX The sequence represents the amino acid sequence of Lfn-Bcl-XL apoptosis-
CC modifying fusion protein comprising anthrax lethal factor (LF) sequence
CC fused to Bcl-XL. The functional apoptosis-modifying fusion protein is
CC capable of binding a target cell and integrating into or crossing a
CC cellular membrane of the target cell. The apoptosis-modifying fusion
CC protein comprises at least two domains: the DFR domain, which targets
CC the fusion protein to the target cell and the Bcl-XL domain, which
CC modifies an apoptotic response of the target cell. The fusion protein is
CC useful for modifying (inhibiting or enhancing) apoptosis in a target
CC cell, such as neuron, lymphocyte, cancer, neoplasm, macrophage,
CC epithelial, stem, tumour or hyper-proliferative cell or an adipocyte. It
CC is also useful for reducing apoptosis in a subject after transient
CC ischaemic neuronal injury, especially spinal cord injury. The fusion
CC protein may be used to treat various diseases and injury conditions
CC through inhibition or enhancement of apoptotic cellular response,
CC including neurodegenerative disorders such as Alzheimer's disease,
CC Huntington's disease, spinal muscular atrophy, stroke episodes and
CC unregulated cell growth as in tumours and various cancers. The apoptosis-
CC modifying fusion protein can be delivered effectively throughout the body
CC and targeted to selective tissue and cells.
XX
XX Sequence 485 AA;
SQ

Query Match 91.3%; Score 1282; DB 22; Length 485;
Best Local Similarity 100.0%; Pred. No. 1.7e-93;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGGHDVGAMHVKEKRNKDEENKRDDEENKTOEHLKEIMKHIVKTEVGEAVKKEAAE 85
DB 22 agghdvgamhvkekeknkdenkrkdeernktgeehkeltmkhivklevgeevkkeaee 81

```
QY 86 KLEKVPDVLKEMTKAIGKTIYVDGDTTKHISLEALSEDKKIKIDYKDALHHEHYV 145
      |||
Db 82 KLEKVPDVLKEMTKAIGKTIYVDGDTTKHISLEALSEDKKIKIDYKDALHHEHYV 141
QY 146 AKEGEYEVLYIQSSEDEVNTEKALNYYEIGKILSRDIISKINOPYOKFLDVNTIKNA 205
      |||
Db 142 AKEGEYEVLYIQSSEDEVNTEKALNYYEIGKILSRDIISKINOPYOKFLDVNTIKNA 201
QY 206 SDSGDGLLFTNOLKEHPPTDFSVFELQNSNEVOEFAKAFAYYIEPOHRDVLQLYAPEA 265
      |||
Db 202 sdsdggdlLftnqlkehpdtfsveflqnsnevegfakafayyiepqhrdvLqlyapea 261
QY 266 FNYMDKFE 274
      |||
Db 262 fnymdkfe 270

RESULT 5
AAR60182
ID AAR60182 standard; Protein; 508 AA.
AC AAR60182;
XX
DT 04-APR-1995 (first entry)
XX
DE LF(1-254)-TR-PE(362-613) toxin fusion protein.
XX
KM Anthrax; Bacillus anthracis; fusion protein; lethal factor;
KM protective antigen; cell killing; targeting; targeting; pathogen;
KM intracellular; HIV; human immunodeficiency virus; toxin;
KM Pseudomonas; exotoxin.
XX
OS Bacillus anthracis.
OS Pseudomonas sp.
XX
PN MO9418332-A.
XX
PD 18-AUG-1994.
XX
PF 14-FEB-1994; 94WO-US01624.
XX
PR 12-FEB-1993; 93US-0021601.
PR 25-JUN-1993; 93US-0082849.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Arora N, Kilmpel K, Leppia SH, Nichols PJ, Singh Y;
XX
DR WPI; 1994-279753/34.
DR N-PSDB; AAO70183.
XX
PT Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumour cells
PT or HIV-infected cells
XX
PS Example 1; Page 95-96; 124pp; English.
XX
CC This sequence is a fusion protein comprising amino acid residues
CC 1-254 of the anthrax protective antigen binding domain of the native
CC anthrax lethal factor, a two residue linker and residues 362-613 of
CC a Pseudomonas exotoxin A activity inducing domain of a second
CC protein. Such fusion proteins may be useful for the specific
CC killing of tumour cells or the killing of cells infected with
CC intracellular pathogens, especially HIV, depending on their
CC components.
XX
SQ Sequence 508 AA;
```

Query Match 91.3%; Score 1282; DB 15; Length 508;
Best Local Similarity 100.0%; Pred. No. 1.8e-93;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 26 AGHGDVGMHYKEKKNKDEENRKTQEEHLKEIMKHIVKIEVGEEAVKKEAAE 85
      |||
Db 1 agpbgdvgmhykexeknkdenkrktqeehlkeimkhivkievgeavkkeaee 60
QY 86 KLEKVPDVLKEMTKAIGKTIYVDGDTTKHISLEALSEDKKIKIDYKDALHHEHYV 145
      |||
Db 61 KLEKVPDVLKEMTKAIGKTIYVDGDTTKHISLEALSEDKKIKIDYKDALHHEHYV 120
QY 146 AKEGEYEVLYIQSSEDEVNTEKALNYYEIGKILSRDIISKINOPYOKFLDVNTIKNA 205
      |||
Db 121 AKEGEYEVLYIQSSEDEVNTEKALNYYEIGKILSRDIISKINOPYOKFLDVNTIKNA 180
QY 206 SDSGDGLLFTNOLKEHPPTDFSVFELQNSNEVOEFAKAFAYYIEPOHRDVLQLYAPEA 265
      |||
Db 181 sdsdggdlLftnqlkehpdtfsveflqnsnevegfakafayyiepqhrdvLqlyapea 240
QY 266 FNYMDKFE 274
      |||
Db 241 fnymdkfe 249

RESULT 6
AAR60178
ID AAR60178 standard; Protein; 776 AA.
AC AAR60178;
XX
DT 03-APR-1995 (first entry)
XX
DE Lethal factor of Bacillus anthracis.
XX
KM Anthrax; Bacillus anthracis; fusion protein; lethal factor;
KM protective antigen; cell killing; targeting; targeting; pathogen;
KM intracellular; HIV; human immunodeficiency virus; toxin.
XX
OS Bacillus anthracis.
OS Pseudomonas sp.
XX
PN MO9418332-A.
XX
PD 18-AUG-1994.
XX
PF 14-FEB-1994; 94WO-US01624.
XX
PR 12-FEB-1993; 93US-0021601.
PR 25-JUN-1993; 93US-0082849.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Arora N, Kilmpel K, Leppia SH, Nichols PJ, Singh Y;
XX
DR WPI; 1994-279753/34.
DR N-PSDB; AAO70179.
XX
PT Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumour cells
PT or HIV-infected cells
XX
PS Disclosure; Page 75-77; 124pp; English.
XX
CC The sequence encoding the lethal factor of Bacillus anthracis may be
CC used in the construction of a nucleic acid which encodes a fusion
CC protein comprising the anthrax protective antigen binding domain of
CC the native anthrax lethal factor and a sequence encoding an activity
CC inducing domain of a second protein. The fusion proteins are useful
CC for the specific killing of tumour cells or the killing of cells
CC infected with intracellular pathogens, especially HIV.
XX
SQ Sequence 776 AA;
```

Query Match 91.3%; Score 1282; DB 15; Length 776;
Best Local Similarity 100.0%; Pred. No. 3.1e-93;

	Matches	249,	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	26	AGGHDDVGMHYKKEKKKNNDENKRKRDEERKKTDEEHLEKMTIIVKIEVGGEEAVKKEAAE	85							
Db	1	agghddvdmhykekknndenkrkdeernktbeehlekmtkhlivievgeeeavkkaeae	60							
Qy	86	KLLEKVPSPDVLMEYFAIGKIIYVGDITRKHISLEALSDKKRKIKDIYKQDALLHEHY	145							
Db	61	kllekvpdvlemyaigkiiyivgdltkhslealsedkkkikdiygkaallheny	120							
Qy	146	AKEGEIEPLVIVQSSDDYVENVTEKALNVYEIEKIIISRLSLKINOPYQKFLDVLNTIKKA	205							
Db	121	akegeieplvivqssddyventekalnvyeyekiiisrllsklnpbyqkflvlnltikna	180							
Qy	206	SDSDQDILLFTNQLKEHPTDSEVFLELEONSNSNQVEFAFAFAFYIIEPQRDVIQIATYAPPA	265							
Db	181	sdsdqdllftnqlkehpdtsevfleleqnsnevgvfaafayilepqrdrvliqliyapea	240							
Qy	266	FNYMDKENE	274							
Db	241	fnymckine	249							

[illegible]

Query Match	30.4%	Score 427;	DB 11;	Length 800;
Best Local Similarity	35.7%	Pred. No. 1.4e-25;		
Matches 100; Conservative	54;	Mismatches 100;	Indels 26;	Gaps 6;

```

QY 2 VISMSCLTATLISGPVPIPLVIGAGGHDGVGMHYKKNKKDEKKRDEERKNTOEHL 61
Db 13 iisfvallfaissqalevna-----mneytsedlkhnhkexnkhkekf 59
QY 62 KEIMKHIVKIEKGEEAAVKAEAKLEKVPDYLEMYKAIGKRIYIVDGDITHISLEA 123
Db 60 kdslnmlvqfethetcdtkigtqtdlkkipkdvleiysejygfcdalvhekeiqd 110
QY 122 LSEPKKTKKDLYGMDALLHENHYVAKBEYEPVLYIGQSESDVEMTEKLNLYNYEPHGKLS 180
Db 120 lseeknsmsmrgekvvpfastrfvekkretcpklll-nkdyplnsegskeyeyelgyls 170
QY 182 RDLISKINQPYQKFLD--VLNFTIKNAS--DSDGDLLFTFTNOUKE---HPTDFSEVLEEQN 234
Db 179 lyiisk-----kdsldpeflnllksldsdssdsdlfsgkfkexkelnlneksldlnfiken 233
QY 235 SNEQVEYFAKAFAYIEHQHHDVYQQLAPAEAFNMMDKNE 274
Db 234 ltefqhalslafsyfapdhvtvleiyapdmfeymunkle 273

```

RESULT	8
AA18294	
ID	AA18294 standard; Protein: 980 AA.
XX	
AC	AA18294;
XX	
DT	07-NOV-2000 (first entry)

DE Plasmodium falciparum chromosome 2 related protein seq ID NO:152.
KM Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
XX Plasmodium falciparum.
OS
XX
XX WO200025728-A2.
PN
XX
XX 11-MAY-2000.
PP
XX
XX 05-NOV-1999; 99WO-US26796.
XX
XX 05-NOV-1998; 98US-0107131.
XX
XX (HOEF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
XX WPI: 2000-365347/31.
DR
XX
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
XX
PS Disclosure; Page 362-365; 577pp; English.

XX AAY20047;
AC
XX
DT 19-JUL-1999 (first entry)
DX
DE B. burgdorferi antigenic protein, t352_aa.
XX
KM Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS Borrelia burgdorferi.
XX
PN MO9859071-A1.
PD 30-DEC-1998.
XX
PE 18-JUN-1998; 98WO-US127718.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
XX
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
DR WPI: 1999-189980/16.
DR N-PSDB; AAX61744.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX
PS Claim 12; Page 172; 275pp; English.
XX
CC This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
SQ Sequence 481 AA;

Query Match 8.7%; Score 122; DB 20; Length 481;
Best Local Similarity 21.7%; Pred. No. 0.1; Matches 51; Conservative 49; Mismatches 81; Indels 54; Gaps 8.

QY 33 GMHVAKKKNDK---NKRKDEERRKTQEHLEKEIKMTNKIVKI-----EYKGEBA 78
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 57 gkdllipenitdkktpqkrmdendlkvsylenyknkkleklktknqtksenekkkies 116
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 79 VKKEEA--EKLEKYPSDYLEMYKAIGKITIYVDGDTITKHISLEALSDKKIKINDYGKD 136
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 117 lekkaekkyelllnkneiveiklknkikpkedenyeklnleniee-----td 167
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 137 ALLHHVYYAKR-----GYEPVLVIQSSEEDYVENTERALKM---VYVIGKILSNDILS 186
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 168 ddfelnyendelextnedy-----psnegllnnlkenlnenekyyalneklddeled 221
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 187 KINQYQKFLDLNTKKNASDSGDGLFTTNNLKEHPDPFSVFLEPONSNEVOEV 241
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 222 rinementlldqrelnrfkkdn-----sknneeleeenlssigrf 263
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:

RESULT 13
ID AAY20046
XX
XX AAY20046;
XX

xx		19-JUL-1999	(first entry)	
xx	B.	burgdorferi	antigenic protein, f352_aa.	
xx	Antigenic	protein; vaccine;	Lyme disease; infection;	detection.
xx	Borrelia burgdorferi.			
xx	MO9859071-A1.			
xx	30-DEC-1998.			
xx	18-JUN-1998;	98WO-US12718.		
xx	03-SEP-1997;	97US-0057483.		
xx	20-JUN-1997;	97US-0050359.		
xx	22-JUL-1997;	97US-0053344.		
xx	22-JUL-1997;	97US-0053377.		
xx	(HUMA-) HUMAN GENOME SCI INC.			
xx	(MEDI-) MEDIMUNE INC.			
xx	Choi GH, Erwin AL, Hanson MS,	Lathigra R;		
xx	WPI: 1999-189980/16.			
xx	N-PSDB; AAX61743.			
xx	New isolated Borrelia burgdorferi nucleic acids - used to develop			
xx	products for the diagnosis, prevention and treatment of diseases			
xx	caused by Borrelia, particularly Lyme disease			
xx	Claim 12; Page 172; 275pp; English.			
xx	This sequence represents a Borrelia burgdorferi (Bb) protein of the			
xx	invention, which is suitable for use in a vaccine. The Bb polypeptides			
xx	can be used in vaccines for eliciting protective antibodies to members of			
xx	the Borrelia genus, particularly for the use against Lyme disease in			
xx	humans and animals. They can be used for preventing or attenuating an			
xx	infection caused by a member of the Borrelia genus. The products can also			
xx	be used for detection of members of the Borrelia genus.			
SQ	Sequence	497 AA;		
Query Match		8.7%; Score 122; DB 20; Length 497;		
Best Local Similarity		21.7%; Pred. No. 0.11;		
Matches	51; Conservative	49; Mismatches	81; Indels	54; Gaps
OY	33 GMHVKKKKNDE--NKRKDEERKKTQEBHLKETLMAKIIVKI-----EYKGGEA	78		
Db	I : I : I : I : I : I : I : I : I : I : I : I : I : I :			
Db	73 gkdIlgpndtkkIprqkrmdendIksvIenyenkiInleklktngktseHenkkies	132		
OY	79 VKKEAA--EKLEKRPDVLLEMUKAIGSKIYIVDSGITKHISLEASDKKKIKNDIYGKD	136		
Db	: :			
OY	137 ALNHVUYAKE-----GYEPVLVIQSSEDPYEVTFKALN---VUYEIGKITLSRDILS	186		
Db	: : I :			
OY	184 ddfenIyendeIextnedny-----psnegilmlkenlneneKyualnekkIdelcd	237		
Db	: : I :			
OY	187 KINPQYKFEDLVNTIKNASDSDGDLFTNQLEKHPDFSVFLEQNSNEVOEV	241		
Db	: : I :			
OY	238 rinementlIdgreflnfkKkdn-----sdknleeIeenlssigrI	279		
RESULT	14			
ID	AAR10534			
XX	AAR10534 standard; Protein:	1427 AA.		
XX	AAR10534;			
DT	12-APR-1991 (first entry)			

XX Human 160kd mediator of inflammation protein.
DE Mediator of Imflammation; cytokine; Hodgkin's lymphoma; MRP-160.
KM Homo sapiens.
OS
XX EP412050-A.
PN
XX 06-FEB-1991.
PD
XX 26-JUN-1990; 90EP-0810481.
PF
XX 05-JUL-1989; 89GB-0015414.
PR
XX (CIBA) CIBA GEIGY AG.
PA
XX Odink KG, Tarcsay L, Bruggen J, Wiesendanger W, Cerletti N;
PI Sorz C, Dewolf-Peters C, Delabie J;
XX WPI: 1991-038913/06.
DR N-PSDB; AA010378.
XX
XX 160 kD human polypeptide mediator or precursor of inflammation
PT polyclonal or monoclonal antibodies to polypeptide treat and
PT diagnose chronic inflammation and hodgkins' lymphoma
XX
XX Claim 3; Page 32; 47pp; English.
PS
XX
XX The protein is a cytokine used to treat chronic inflammatory
CC conditions. It is prepared by chromatographically purifying an
CC optionally pre-purified cell extract, cell supernatant or cell
CC filtrate of stimulated normal human leucocytes or human embryonic
CC epithelial lung cells. Alternatively, the protein can be produced
CC by microorganisms or continuous mammalian cell lines, transformed
CC with plasmids encoding MRP-160. The invention also covers the
CC polypeptide fragment from amino acids 878-1427 and derivatives of
CC the protein in which the amino and/or hydroxyl functions are
CC glycosylated or acylated and have mol. wt.s of 190 and 140kD,
CC respectively.
XX
XX Sequence 1427 AA;
SQ

Query Match 8.5%; Score 119; DB 12; Length 1427;
Best Local Similarity 21.7%; Pred. No. 0.7;
Matches 60; Conservative 51; Mismatches 99; Indels 66; Gaps 13;

OY 36 VKEKKNKDNKRR-----DEBRNKTOEHLKEIMKHIVKIEVGGEPAVK--- 80
DB 692 lkeKensleairskIdkeedqhlvemedtlnkIgeaek--vkelevIdakcneqtkynd 749
OY 81 -----KEAAEKLE-----KVPSPVLEMYKAIGKRIYVDGDKHISLELSEDPK-- 127
DB 750 nftsglkateekllldldalrkassegsemkklrgleaaeqi-khlelekaesskas 808
OY 128 -KINDIYVKDALH-----EHYVAKGEYEPVLIQSSSEDEVNTERKALNYYIEIGKL 180
DB 809 snrtelgtrrekltnlgnelsevsgvketlekeql-i-kektaaeseeavsvgrism---- 863
OY 181 SRDLSKINOPKFLVDVLTNIKNASDSOGDLFTNOLKEHPFPFSVEFLQNSNEVOE 240
DB 864 -getvnrklnhgkeqf-----lmalsd-----leklrenladmeakfrefdereqj 908
OY 241 VFAAFAYIEPQHRDVLQVLAPEAFN--YMDKFE 274
DB 909 lkak-----eklendlaeimkgsgdnassqtlkmd 938

RESULT 15
AAW18010
ID AAW18010 standard; Protein: 652 AA.
XX

AC AAW18010;
XX
XX 08-DEC-1997 (first entry)
DT
XX Plasmodium falciparum sequestrin.
DE
XX Plasmodium falciparum sequestrin.
KM Sequestrin; malaria; vaccine; diagnosis.
XX
XX Plasmodium falciparum clone 3D7.
OS
XX
XX Key
FH
XX Peptide
FT 9..18 Location/Qualifiers
FT /note= "Asn/Thr-rich decamer repeat"
FT 19..28
FT Peptide
FT /note= "Asn/Thr-rich decamer repeat"
FT 29..38
FT Peptide
FT /note= "Asn/Thr-rich decamer repeat"
FT 42..180
FT Domain
FT /label= "CD36-binding_domain"
FT /note= "(Claim 24)"
FT
FT Misc-difference
FT 111
FT /note= "encoded by GAA"
FT
FT Misc-difference
FT 119
FT /note= "encoded by GAA"
FT
FT Misc-difference
FT 120
FT /note= "encoded by GAA"
FT
FT Misc-difference
FT 340
FT /note= "encoded by CAA"
FT
FT Peptide
FT 542..547
FT /note= "Asn/Asp-rich hexamer repeat"
FT 548..553
FT Peptide
FT /note= "Asn/Asp-rich hexamer repeat"
FT 554..559
FT Peptide
FT /note= "Asn/Asp-rich hexamer repeat"
FT 560..565
FT Peptide
FT /note= "Asn/Asp-rich hexamer repeat"
FT 566..571
FT Peptide
FT /note= "Asn/Asp-rich hexamer repeat"
FT 572..577
FT Peptide
FT /note= "Asn/Asp-rich hexamer repeat"
FT 578..583
FT Peptide
FT /note= "Asn/Asp-rich hexamer repeat"
FT 584..589
FT Peptide
FT /note= "Asn/Asp-rich hexamer repeat"
FT 590..595
FT Peptide
FT /note= "Asn/Asp-rich hexamer repeat"
FT 596..601
FT Peptide
FT /note= "Asn/Asp-rich hexamer repeat"
FT 602..607
FT Peptide
FT /note= "Asn/Asp-rich hexamer repeat"
FT 608..613
FT Peptide
FT /note= "Asn/Asp-rich hexamer repeat"
FT 614..619
FT Peptide
FT /note= "Asn/Asp-rich hexamer repeat"
FT 620..625
FT Peptide
FT /note= "Asn/Asp-rich hexamer repeat"
FT

W09719168-A1.
PD 29-MAY-1997.
XX
XX 15-NOV-1996; 96WO-US17265.
PF
XX 17-NOV-1995; 95US-0559896.
PR
XX
XX (USSA) USARMRC US ARMY MEDICAL RES MATERIAL COMMAND.
PA Duffy PE, Ockenhouse CF;
XX
XX WPI: 1997-298107/27.
DR N-PSDB; AAT67161.
XX
XX New DNA coding for sequestrin protein from Plasmodium falciparum -

PT used for vaccines against, and diagnosis of, malaria

XX Claim 23; Page 37-39; 61pp; English.

CC This polypeptide comprises sequesterin, a cytoadherence protein
 CC specific to Plasmodium falciparum which binds to CD36 and is
 CC involved in sequestration of P. falciparum in the deep vascular
 CC beds of various tissues. Isolated, purified sequesterin, or its
 CC CD36-binding domain (see AAW18011), can be expressed in transformed
 CC prokaryotic or eukaryotic host cells using claimed vectors
 CC incorporating sequesterin DNA (see AAT67161-62). A claimed vaccine
 CC against malaria contains sequesterin protein, and a claimed
 CC malaria diagnostic kit contains antibody raised against sequesterin.
 CC Sequesterin can also be used in a claimed method to test for agents
 CC capable of inhibiting sequestration. Such agents, as well as anti-
 CC sequesterin antibodies, can be administered to a patient in a
 CC claimed method for preventing or ameliorating symptoms of malaria
 CC infection.

XX
 SQ Sequence 652 AA;

Query Match 8.4%; Score 118.5; DB 18; Length 652;

Best Local Similarity 23.1%; Pred. No. 0.28; Mismatches 54; Conservative 57; Indels 51; Gaps 11;

QY 38 EKEKNKDEENKRRKDEERNKTOEHL-----KETMKHIVKIEVKGEPAYKKEAAEKLEK 90
 Db 197 eketnkdknkdkidididvdidkdhve-elyge--vknklskeeldr 253
 QY 91 VPSDVL-----EMYKAIGKITYVDGDTKHSLEALSE-DKKRTKDIYGR 135
 Db 254 mtdaalyrvyleeldrmdelrvyleelekdkeekelhreklhlekinkm-dk 312
 QY 136 DALLHEHYVYAKEGEPVLVYOSSEDEVVENTEKALNVYEIGKILSRDILSKINQPYQKF 195
 Db 313 dql---dklyeee-----lnkmsdeiqhvrriyledikelqnle-leeidrllykee 362
 QY 196 LDVIN-----TIKNASDSGDGLFTNOLKEHPTDFSVFLEONSNEVOEVEF 242
 Db 363 ldrndrearyelpmrnlrnsneknlnhn-----lnesngknkknenvvf 408

Search completed: December 2, 2001, 13:48:22
 Job time: 142 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: December 2, 2001, 13:51:02 ; Search time 72.22 Seconds
(without alignments)
289.003 Million cell updates/sec

Title: US-09-747-521-2_COPY_9_282
1404
Sequence: 1 KVISMSCLVTAITLSCGVFI.....RDVLQLYAPFAFNMDKPFNE 274

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1404	100.0	809	1	J00032	anthrax toxin leth
2	442	31.5	800	1	U50029	adenylate cyclase
3	442	31.5	800	2	B59106	hypothetical prote
4	135.5	9.7	1408	2	T45039	hypothetical prote
5	129.5	9.2	583	2	A70380	arginine-tRNA lig
6	129.5	9.2	980	2	E71606	hypothetical prote
7	129	9.2	1245	2	D71613	GAF domain protein
8	125.5	8.9	1526	2	A45605	mature-parasite-in
9	125	8.9	558	2	C71609	hypothetical prote
10	124	8.8	649	2	T27232	hypothetical prote
11	123.5	8.8	880	2	F75103	conserved hypotet
12	122.5	8.7	508	2	E71620	hypothetical prote
13	122	8.7	497	2	H70168	hypothetical prote
14	122	8.7	1225	2	A49464	chromosome segrega
15	122	8.7	1624	2	T25592	hypothetical prote
16	121	8.6	2829	2	A42771	reticulocyte-bindi
17	120.5	8.6	1156	2	B70356	chromosome assembl
18	120	8.5	709	1	HHBY90	heat shock protein
19	119	8.5	2269	1	T28677	rhocty protein -
20	118.5	8.4	378	2	T18486	hypothetical prote
21	118	8.4	1005	2	A64465	microtubule-vesicl
22	118	8.4	1392	2	A43336	restin - human
23	118	8.4	1427	2	S22695	rhocty protein -
24	118	8.4	2401	2	T28676	plis homolog
25	117	8.3	1169	2	A64505	hypothetical prote
26	117	8.3	1351	2	C71607	hypothetical prote
27	116.5	8.3	652	2	T41162	recombination prot
28	116	8.3	520	2	F70350	heat shock protein
29	116	8.3	705	2	S55133	

30	116	8.3	1087	2	T30330	gelsoiin-related p
31	115.5	8.2	886	2	H69378	conserved hypotet
32	115.5	8.2	962	1	S57342	endopeptidase La h
33	114.5	8.2	284	2	S24401	tropomyosin 2, ske
34	114.5	8.2	1365	2	T45031	hypothetical prote
35	113.5	8.1	3351	2	T13812	lipophorin - fruit
36	113	8.0	435	2	A86492	hypothetical prote
37	113	8.0	435	2	H72129	hypothetical prote
38	113	8.0	593	2	S27483	transferrin-bindin
39	113	8.0	593	2	B44796	transferrin-bindin
40	113	8.0	605	2	S46833	hypothetical prote
41	113	8.0	805	2	E70474	translation initia
42	113	8.0	975	1	A31497	kinesin heavy chai
43	113	8.0	1280	2	G96796	hypothetical prote
44	113	8.0	1979	2	C71622	hypothetical prote
45	113	8.0	2231	2	D71870	hypothetical prote

ALIGNMENTS

RESULT 1
J00032
anthrax toxin lethal factor pXOI-107 precursor - Bacillus anthracis virulence plasmid
C:Species: Bacillus anthracis
C:Date: 31-Mar-1990 #sequence_revision 11-Nov-1994 #text_change 11-May-2000
C:Accession: J00032; C59104
R:Bragg, T.S.; Robertson, D.L.
Gene 8L, 45-54, 1989
A:Title: Nucleotide sequence and analysis of the lethal factor gene (lef) from Bacill
A:Reference number: J00032; M0ID:90034185
A:Accession: J00032
A:Molecule type: DNA
A:Residues: 1-809

A:Cross-references: GB:M29081; NID:g143143; PIDN:AAV9216.1; PID:g143144
R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koeh
J: Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pXOI, the large Bacillus anthracis plasmid harb
A:Reference number: A59091; M0ID:99445483
A:Accession: C59104
A:Molecule type: DNA
A:Residues: 1-809 <OKI>
A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32411.1; PID:g4894323
A:Experimental source: strein Sterne
A:Note: Similar to Anthrax toxin lethal factor precursor. lef, plasmid pXOI, B. anthr
C:Comment: This lethal factor of Bacillus anthracis is part of the tripartite protein
her they cause anthrax, an infectious and often fatal disease of cattle, sheep, and o
C:Genetics:
A:Gene: lef; pXOI-107
A:Genome: plasmid
C:Superfamily: anthrax toxin lethal factor; lethal factor amino-terminal homology
C:Keywords: toxin
F:1-33/Domain: signal sequence #status predicted <SIG>
F:34-809/Product: anthrax toxin lethal factor #status predicted <MAY>
F:44-295/Domain: lethal factor amino-terminal homology <LFA>

Query Match 100.0%; Score 1404; DB 1; Length 809;
Best Local Similarity 100.0%; Pred. No. 3.2e-72;
Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVISMSCLVTAITLSCGVFIPLVQAGSGHDVGMHVKENKNDENRKDEERKQDEH 60
DB 9 KVISMSCLVTAITLSCGVFIPLVQAGSGHDVGMHVKENKNDENRKDEERKQDEH 68
QY 61 LKTIKMHIVIEVKGEEAVKKEAEKLEKVPDVLMEYKAIGKITYIVDGDTKKHSLE 120
DB 69 LKTIKMHIVIEVKGEEAVKKEAEKLEKVPDVLMEYKAIGKITYIVDGDTKKHSLE 128
QY 121 ALSEDEKKIKDIYKDALHEHYVYAKEGYEPVLVIQSSSEDEYVENTEKALNYYEIGKIL 180
DB 129 ALSEDEKKIKDIYKDALHEHYVYAKEGYEPVLVIQSSSEDEYVENTEKALNYYEIGKIL 188

[illegible]

RESULT 11
F75103
conserved hypothetical protein PAB0812 - *Pyrococcus abyssi* (strain Orsay)
C:Species: *Pyrococcus abyssi*
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 02-Feb-2001
C:Accession: F75103
R:Anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome structure
A:Reference number: A75001
A:Accession: F75103
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-880 <RAW>
A:Cross-references: GB:AJ248286; GB:AL096836; NID:g5458366; PIDN:CAP50131.1; PID:g5458666
C:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB0812
A:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

```

Query Match      8.8% Score 123.5; DB 2; Length 880;
Best Local Similarity 23.3%; Pred. No. 9.7;
Matches 70; Conservative 50; Mismatches 94; Indels 87; Gaps 15;

OY      31 DVGIMVKEKKEKKDKENKKRDEERNKQ-----EEHKLEIKMKHIVKIEVG-EEA 78
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      413 EITTTIGMEDEKENRMKAIEELRKAKGCPYCGRELTBEHKELMERT-LEIKKIEE 471
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY      79 VKKEAAE-----KLE-KVPSDVLEMYKAIGKI-YIVDDITKHISLEAL 122
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      472 LKRTTEERKRLVNLKRLKLEIKLREPSVMRDIAEQIKLEESKLGKGNLELEQKEREFEGL 531
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY      123 SE-----DKKKIKDIYKQALLHENVYAKE-----GYEPV-- 153
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      532 NEEFNKLKGELLGERLDERKRIKALGRRKLIIEKYRKAKEELENLHRLQRELGFEFSVEEL 591
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY      154 -LVIOSSD---YVENTEKALNVEYEGLKILSRILSKINPYOKFLDVLTTINASDS 208
      | | | | : | : : : : : : : : : : : : : : : : : : : : : :
Db      592 NLRIOLEEFHDKYVE-AKKSSELRLELKNKKLEKE-KTELDAFELADAVENEIEEKE-- 647
      | | | | : | : : : : : : : : : : : : : : : : : : : : : :

OY      209 DGODLLFTNOLKEHPTDSEVFLEQN-----SNEVOEFAKAFAYVIEPOHRLVLOLY 261
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      648 -----AKLKDELSKINEEYEEKRERVLKLENEVSSLTAR-----LEELKKSVEQIK 694
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY      262 A 262
      |
Db      695 A 695

```

hypothetical protein PF00225c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: E71620
R:Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aarand, L.; Koonin, E.; Perles, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: E71620
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-508 <GAR>
A:Cross-references: GB:AE001381; GB:AE001362; NID:93845124; PIDN:AMC71834.1; PID:g384
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PF00225c

Query Match	8.7%	Score 122.5;	DB 2;	Length 508;
Best Local Similarity	28.4%;	Pred. No. 5.7;		
Matches	48;	Conservative	32;	Mismatches 42; Indels 47; Gaps 10;
QY	37	KEKEKN----	KDNKRKQDEERNKQDEHLKELMTMHIYIEVKGGEPAVKKEAEKLLKVP	92
		:	: : : :	:
DB	376	KEKOKNKKNEKKNKKKEKEKKNKKKEKE--	KKKKKKEKESKKKEKEKKNKKKKR--	432
QY	93	SDVLEMYKAIGKIITVGDITKHISLEASBDDKKIKIIDIYQKALLLHEHVYAKGEYR		152
			: :	: :
DB	433	-----EKEKN-----	NGDVLKHH-----ENNLQDV-----ELLYE-----EK	459

QY 153 VVIOSSSEYEYENTKALNYEIKLISRLPSKINO--PYOGFELVL 199
:::---|::| | | | | | | | | | |
DB 460 IIVNVTKRDELSLK---NKYS-KDIVHILSEISNTLYTSFLDYM 503

RESULT 13
H70168
hypothetical protein BB0553 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: H70168
R:Fischer, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kurlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A>Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: AV0100; MUID:98065943
A:Accession: H70168
A>Status: preliminary. nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1 497 <KE>
A:Cross-references: GB:AE001157; GB:AE000783; NID:92688471; PIDN:AAC66928.1; PID:g268
A:Experimental source: strain B31

	Query Match	8.7%	Score 122;	DB 2;	Length 497;	
	Best Local Similarity	21.7%;	Pred. No. 6;			
	Matches	51; Conservative	49;	Mismatches	81;	Indels 54; Gaps 8;
QY	33 GMHVEKEKNKOE--NKRROEERKNTJOEHKLKIMKHIVKI-----EVKGEEA	78				
	: : : : : : : : : : : : : : : :					
Db	73 GMDLPIRIRKKLPQRKDENDDKLSVYLEENKIKINIEKLLTKRNOKTSSENKKIIS	132				
QY	--VKREA--EKLEKVPSDYLEMYKAIGKTIYVDGDTFKHSLEALSSEDKKKIKIDYGKD	136				
	: : : : : : : : : : : : : : : : : : : : : : : : : :					
Db	133 IRRKAKKYEILNLKLNELIVEIKLLNKRIKPKDEBNYEKINIENEED-----TD	183				
QY	137 ALLHEHYIAKE-----GYEPVLVIOSSEDYVENTEALN---YYEIGKITLSRLIS	186				
	: : : : : : : : : : : : : : : :					
Db	184 DDFEENFEYNDEIXNETDNY-----PSNEOIIINLKLENENEKYVAINKKIDELED	237				

QY 187 KINPYOKFLDVLWTINKNASDSDGODLLFTNOLKEHPDSEVFELEONSNEVOEY 241
 DB 238 RINEMENTIIDLQRELFNFKKKN-----SDKNLEEFIEENLSSIGRI 279

RESULT 14

A49464

Chromosome segregation protein SMCI - yeast (*Saccharomyces cerevisiae*)

N/Alternate names: probable head-rod-tail protein SMCI; protein YFL008W

C/Species: *Saccharomyces cerevisiae*

C/Date: 03-May-1994 #sequence-revision 03-May-1994 #text-change 20-Jun-2000

C/Accession: A49464; S56246; S62305; S41804

R:Strunnikov, A.V.; Larionov, V.L.; Koshland, D.

J. Cell Biol. 123, 1635-1648, 1993

A>Title: SMCI: an essential yeast gene encoding a putative head-rod-tail protein is requ

A/Reference number: A49464; MUID:94103320

A/Accession: A49464

A/Molecule type: DNA

A/Residues: 1-1225 <STR>

A/Cross-references: GB:L00602; NID:g172620; PIDN:AAA16595.1; PID:g172621

R:Murakami, Y.; Naitou, M.; Hagihara, H.; Shibata, T.; Ozawa, M.; Sasamura, S.I.; Sasana

submitted to the EMBL Data Library, May 1995

A/Description: Analysis of the nucleotide sequence of chromosome VI from *Saccharomyces ce*

A/Reference number: S56186

A/Accession: S56246

A/Molecule type: DNA

A/Residues: 1-1225 <MUR>

A/Cross-references: EMBL:D50617; NID:g83685; PIDN:BA09230.1; PID:g836746; MIPS:YFL008W

R:Murakami, Y.

submitted to the EMBL Data Library, April 1994

A/Reference number: S62302

A/Accession: S62305

A/Molecule type: DNA

A/Residues: 1-1225 <MUR>

A/Cross-references: EMBL:D31600; NID:g836814; PIDN:BA06496.1; PID:g836816

C/Genetics:

A:Gene: SGD:SMCI

A/Cross-references: SGD:S0001886; MIPS:YFL008W

A:Map position: 6L

C/Function:

A:Description: probably involved in chromosome segregation

C:Superfamily: chromosome segregation protein SMCI

C:Keywords: cell division; coiled coil; mitosis; nucleus

Query Match 8.7%; Score 122; DB 2; Length 1225;

Best Local Similarity 21.9%; Pred. No. 18;

Matches 62; Conservative 50; Mismatches 89; Indels 82; Gaps 13;

QY 32 VGMHVKKKKKDKKDEKRNKTQEHLEIKKHIVKIGKE-EAVKKEAEKLEK 90
 DB 796 IGFITKEVENHSGELMRQ-----OSKELOOLQKQILTVENKLOFETDRLSTQRRYEK 848
 QY 91 VPSVULEMYKAIIGKITYIVDGIKHSLEALSDKKIKDIYK-DALLHEHYVYAKG 149
 DB 849 AQKD-LE-----NAQVEMKSLSEDEVAIEKKSIESKLEEHKHNHDEL 891
 QY 150 YEPVLV-----IQSESDYVENTEKALNYYEIGKILSRDIISKINOPYOKF-LDVLNTIKN 204
 DB 892 QKKFVTQKSELNSEDLEDNSNL-----QYLKRE-RGCIKEDIKFLDERVTALKN 943
 QY 205 A-----SDSDGODLLFTNOL-----KEHPDSEVFELEQ 233
 DB 944 CKISINIPISSETTIDLPISSTDNENAITISNIDINVKGLPKRYKENNTDSARKELEQ 1003
 QY 234 NSNEVOEV-----AKAFAYIEPQHR-----DVLQVLAPE 264
 DB 1004 KIHVEEELNLEQPNARALEYDEAGREYVYINNETQLKAE 1046

RESULT 15

T25592

hypothetical protein C32E12.4 - *Caenorhabditis elegans*

C/Species: *Caenorhabditis elegans*
 C/Date: 15-Oct-1999 #sequence-revision 15-Oct-1999 #text-change 15-Oct-1999
 C/Accession: T25592

R:Willcox, L.

submitted to the EMBL Data Library, November 1996

A/Description: The sequence of *C. elegans* cosmid C32E12.

A/Reference number: Z20055

A/Accession: T25592

A>Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1624 <WIL>

A/Cross-references: EMBL:U80032; PIDN:AB53680.1; GSPDB:GN00019; CESP:C32E12.4

A/Experimental source: strain Bristol N2; clone C32E12

A:Gene: CESP:C32E12.4

A:Map position: 1

A/Introns: 13/2; 123/2; 175/1; 249/3; 325/3; 357/3; 513/3; 570/1; 617/3; 674/3; 694/1

Query Match 8.7%; Score 122; DB 2; Length 1624;

Best Local Similarity 25.6%; Pred. No. 25;

Matches 58; Conservative 35; Mismatches 60; Indels 74; Gaps 12;

QY 37 KEKKNKDKNRKDEE-----RNKTQEHLEIKM-----KHIVK---IEYKGEA 78
 DB 1053 KKKQEEEEKKKRKEKEMKKNKEEEKLKRDIEAKSLAAGEDKSALRKSALNMKDEDA 1112
 QY 79 VKKEAEKLEKVPDVLWYKAIIGKITYIVDGIKHSLEALSDKKIKDIYGNAL 138
 DB 1113 KKKVATD---EKVAQAQARRF-----GIYVYAG-----KIDETFEAKK----- 1149
 QY 139 LHEHYVYAKGEYEPVLVQSESDYVENTEKALNYYEIGKILSRDIISKINOPYOKFLD- 197
 DB 1150 -----KE---EKVLYKSSRLFAKNADKPKRYEVYIKPV-----INDDFDKWDE 1191
 QY 198 VLNTIKNASDSDGODLLFTNOLKEHPDSEVFEPL-----EONSNEVOE 240
 DB 1192 IRKQMKSGS-----NQLQSAIKDLSKGLISATEAKSREMEE 1228

Search completed: December 2, 2001, 13:51:05
 Job time: 300 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:51:46 ; Search time 44.8 Seconds

(Without alignments)
224.245 Million cell updates/sec

Title: US-09-747-521-2_COPY_9_282
Perfect score: 1404
Sequence: 1 KVISMSCLVTAITLGGPVFI.....RDVQLVAPFAFNMYDKFNE 274

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1404	100.0	809	LEF_BACAN	P15917 bacillus an
2	442	31.5	800	CYAA_BACAN	P40136 bacillus an
3	129.5	9.2	583	SYR_AOUAE	O67068 aquifex aeo
4	122	8.7	1225	SMCI_YEAST	P32908 saccharomyc
5	121	8.6	2869	RBPI_PLAVB	O00798 plasmodium
6	120	8.5	709	HS82_YEAST	P02829 saccharomyc
7	118	8.4	1427	REST_HUMAN	P30622 homo sapien
8	116	8.3	520	RECN_AOUAE	O66834 aquifex aeo
9	116	8.3	704	HS83_YEAST	P15108 saccharomyc
10	115.5	8.2	318	YZ11_AOUAE	O66405 aquifex aeo
11	114.5	8.2	284	TPME_CHICK	P18441 gallus gall
12	113	8.0	605	APM2_YEAST	P38700 saccharomyc
13	113	8.0	805	IF2_AOUAE	O67825 aquifex aeo
14	113	8.0	975	KINH_DROME	P17210 drosophila
15	112	8.0	1265	DYNA_DROME	O01174 xenopus lae
16	111.5	7.9	248	TPM1_XENLA	P13496 drosophila
17	111.5	7.9	822	FEER_HUMAN	P16591 homo sapien
18	111	7.9	1164	BAG_STRAC	P27991 streptococ
19	111	7.9	2349	TPR_HUMAN	P12270 homo sapien
20	110	7.8	387	YCBR_BACSU	P40758 bacillus su
21	110	7.8	1130	YL17_CAEEL	O11102 caenorhabdi
22	109.5	7.8	548	CEAK_ECOLI	O47502 escherichia
23	109.5	7.8	729	KAR3_YEAST	P17119 saccharomyc
24	109	7.8	373	FLHF_AOUAE	O67266 aquifex aeo
25	109	7.8	1682	MSP1_PLAF3	P19598 plasmodium
26	108.5	7.7	245	TPM5_CHICK	P49439 gallus gall
27	108.5	7.7	281	TPM6_CHICK	P49439 gallus gall
28	108.5	7.7	959	LOM1_HUMAN	P36776 homo sapien
29	107.5	7.7	284	TPM2_CHICK	O04268 gallus gall
30	107.5	7.7	284	TPM3_CHICK	O08992 coturnix co
31	107.5	7.7	284	TPM5_CHICK	P49436 gallus gall
32	107.5	7.7	284	TPM5_COTJA	P49437 coturnix co
33	107	7.6	284	TPM5_COTJA	P02559 coturnix co

ALIGNMENTS

RESULT 1	LEF_BACAN	STANDARD:	PRT:	809 AA.
AC	P15917:			
DT	01-APR-1990 (Rel. 14, Created)			
DT	01-APR-1990 (Rel. 14, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	LETHAL FACTOR PRECURSOR (EC 3.4.24.-) (LF).			
GN	LEF.			
OS	Bacillus anthracis.			
OG	Plasmid pXOI.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_Taxid-1392;			
RN	[1]			
RP	SEQUENCE FROM N.A. AND SEQUENCE OF 34-49.			
RX	MEDLINE=90034185; PubMed=2509294;			
RA	Braeg T.S., Robertson D.L.;			
RT	"Nucleotide sequence and analysis of the lethal factor gene (lef)			
RT	from Bacillus anthracis.";			
RL	Gene=81:45-54(1989).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Lowe J.;			
RL	Submitted (APR-1990) to the EMBL/Genbank/DBJ databases.			
RN	[3]			
RX	MEDLINE=95154669; PubMed=7851740;			
RA	Koch S.K., Schlavo G., Montecucco C.;			
RT	"Zinc content of the Bacillus anthracis lethal factor.";			
RL	(FEMS Microbiol. Lett. 124:343-348(1994).			
CC	-1- FUNCTION:--ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,			
CC	AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE			
CC	DEATH. LEF IS THOUGHT TO BE A LETHAL FACTOR THAT, WHEN ASSOCIATED			
CC	WITH PA, CAUSES DEATH. LEF IS NOT TOXIC BY ITSELF. PA IS THOUGHT TO			
CC	BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC CELLS, THEREBY			
CC	FACILITATING THE INTERNALIZATION OF LEF OR EF.			
CC	-1- SUBUNIT: SECRETED ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT			
CC	PROTEINS. A PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN			
CC	EDERA FACTOR (EF). NONE OF THESE IS TOXIC BY ITSELF.			
CC	-1- SUBCELLULAR LOCATION: SECRETED.			
CC	-1- SIMILARITY: THE PA-BINDING REGION IS FOUND IN BOTH B.ANTHRACIS EF			
CC	AND LEF.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M34 (ZINC			
CC	METALLOPROTEASE).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M29081; AAA79216.1; -			
DR	EMBL; M30210; AAA22569.1; -			

```
DR PIR: J00032: J00032.
DR MEROPS: M34.001: -.
DR InterPro: IPR003541: Anthrax toxinA.
DR InterPro: IPR00130: Zn_MTPeptide.
DR PROSITE: PS00142: ZINC_PROTEASE. 1.
KW Hydrolyase: Metalloprotease; Zinc; Toxin; Signal; Repeat; Plasmid.
FT SIGNAL 1 33
FT CHAIN 34 809 LETHAL FACTOR.
FT DOMAIN 34 293 PA-BINDING REGION (POTENTIAL).
FT DOMAIN 300 420 REPEATS.
FT METAL 719 719 ZINC (CATALYTIC) (POTENTIAL).
FT ACT_SITE 720 720 POTENTIAL.
FT METAL 723 723 ZINC (CATALYTIC) (POTENTIAL).
SQ SEQUENCE 809 AA: 93786 MW: 861684D277310AE CMC64;

Query Match 100.0%; Score 1404; DB 1: Length 809;
Best Local Similarity 100.0%; Pred. No. 17e-71;
Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVISMCLVTAITLSPVPIPLVVGAGHGDMHVKKKKKDKKKDEBNKTOEHH 60
DB 9 KVISMCLVTAITLSPVPIPLVVGAGHGDMHVKKKKKDKKKDEBNKTOEHH 68
QY 61 LKEIMKHIVKIEVGEAEVKKAEKLEKPSDYLEMKYKAGKIIYVDDITKHISLE 120
DB 69 LKEIMKHIVKIEVGEAEVKKAEKLEKPSDYLEMKYKAGKIIYVDDITKHISLE 128
QY 121 ALSEDKKKIKIYGDALHHEHYVAKGEYEVLYQSSQEDVENTEKALNYYEIGKIL 180
DB 129 ALSEDKKKIKIYGDALHHEHYVAKGEYEVLYQSSQEDVENTEKALNYYEIGKIL 188
QY 181 SRDLISKINQYKFLDVNTIKNASDSGDGLFTNOLKEHPDTSVEEFLQNSNEVOE 240
DB 189 SRDLISKINQYKFLDVNTIKNASDSGDGLFTNOLKEHPDTSVEEFLQNSNEVOE 248
QY 241 VFAKFAFYIEPQHRDVLQYAPFAFNMDKPFNE 274
DB 249 VFAKFAFYIEPQHRDVLQYAPFAFNMDKPFNE 282

RESULT 2
CYAA_BACAN
ID CYAA_BACAN STANDARD; PRT: 800 AA.
AC P40136;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CALMODULIN-SENSITIVE ADENYLATE CYCLASE PRECURSOR (EC 4.6.1.1) (ATP
DE PYROPHOSPHATE-LYASE) (ADENYL CYCLASE) (EDEMA FACTOR) (EF).
GN CYA.
OS Bacillus anthracis.
OG Plasmid pXOI.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RN [1] SEQUENCE FROM N.A.
RX MEDLINE=89138004; PubMed=2906312;
RA Escuyer V., Duflot E., Sezer O., Danchin A., Mock M.;
RT "Structural homology between virulence-associated bacterial adenylate
RT cyclases.";
RL Gene 71:293-298(1988).
RN [2]
RN [2] SEQUENCE FROM N.A.
RX MEDLINE=89211974; PubMed=3149607;
RA Robertson D.L., Tippetts M.T., Leppia S.H.;
RT "Nucleotide sequence of the Bacillus anthracis edema factor gene
RT (cya) a calmodulin-dependent adenylate cyclase.";
RL Gene 73:363-371(1988).
RN [3]
RN [3] SEQUENCE FROM N.A.
RA Escuyer V., Duflot E., Mock M., Danchin A.;
```

```
RT "Nucleotide sequences expressing adenylate cyclase from B.anthraxis,
RT proteins having the activity of this adenylate cyclase and biological
RT uses.";
RL Patent number EP0366550, 02-MAY-1990.
RN [4]
RN [4] SEQUENCE FROM N.A.
RX MEDLINE=88198021; PubMed=2834337;
RA Tippetts M.T., Robertson D.L.;
RT "Molecular cloning and expression of the Bacillus anthracis edema
RT factor toxin gene: a calmodulin-dependent adenylate cyclase.";
RN J. Bacteriol. 170:2263-2266(1988).
RN [5]
RN [5] SEQUENCE OF 34-48.
RX MEDLINE=89211974; PubMed=3149607;
RA Robertson D.L., Tippetts M.T., Leppia S.H.;
RL Gene 73:363-371(1988).
RN [6]
RN [6] REVIEW.
RX MEDLINE=93119764; PubMed=8418825;
RA Danchin A.;
RT "Phylogeny of adenylate cyclases.";
RL Adv. Second Messenger Phosphoprotein Res. 27:109-162(1993).
CC -I- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,
CC AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE
CC DEATH. EF IS A CALMODULIN-DEPENDENT ADENYLATE CYCLASE THAT, WHEN
CC ASSOCIATED WITH PA, CAUSES EDEMA. EF IS NOT TOXIC BY ITSELF. PA IS
CC THOUGHT TO BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC CELLS,
CC THEREBY FACILITATING THE INTERNALIZATION OF EF OR LF.
CC -I- CATALYTIC ACTIVITY: ATP -> 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC -I- ENZYME REGULATION: ACTIVATED BY HOST CALMODULIN.
CC -I- SUBUNIT: ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT PROTEINS, A
CC PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN EDEMA FACTOR
CC (EF). NONE OF THESE IS TOXIC BY ITSELF.
CC -I- SIMILARITY: BELONGS TO ADENYLATE CYCLASE CLASS-2 FAMILY.
CC -I- SIMILARITY: THE PA-BINDING REGION IS FOUND IN BOTH B.ANTHRACIS EF
CC AND LF.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL: M23179; AAA22374.1; -.
DR EMBL: M24074; AAA79215.1; -.
DR EMBL: A07289; CAA00652.1; ALT_SEQ.
DR InterPro: IPR003541; Anthrax toxinA.
KW Lyase; CAMP synthesis; Toxin; ATP-binding; Signal; Plasmid.
FT SIGNAL 1 33
FT CHAIN 34 800 CALMODULIN-SENSITIVE ADENYLATE CYCLASE.
FT DOMAIN 34 288 PA-BINDING REGION (POTENTIAL).
FT DOMAIN 289 680 CATALYTIC.
FT NP_BIND 347 354 ATP (POTENTIAL).
FT CONFLICT 350 350 V -> E (IN REF. 2).
FT CONFLICT 510 510 Q -> T (IN REF. 2).
FT CONFLICT 512 513 EM -> RM (IN REF. 2).
FT CONFLICT 760 760 V -> L (IN REF. 3).
SQ SEQUENCE 800 AA: 92477 MW: FAF7EB485DFPC5A6 CMC64;

Query Match 31.5%; Score 442; DB 1: Length 800;
Best Local Similarity 36.4%; Pred. No. 7.8e-18;
Matches 102; Conservative 53; Mismatches 99; Indels 26; Gaps 6;

QY 2 KVISMCLVTAITLSPVPIPLVVGAGHGDMHVKKKKKDKKKDEBNKTOEHL 61
DB 13 IISFVLLFAISSQAEVNA-----MNEHYTESDIKRNHKTKEKKEKF 59
QY 62 KEIMKHIVKIEVGEAEVKKAEKLEKPSDYLEMKYKAGKIIYVDDITKHISLEA 121
```



```

Db      60 KOSINNLVTEFTNETLKDIOQTODLLKPKPDVLEIYSELGEIYFTDIDVHEKELD 119
Qy      122 LSEDDKKIKDIYGRDALLHEHYVYAKEGEFVLYIOSSDDYVENTEKALNYYEIGKIIS 181
Db      120 LSEEEKNSMNSNGEVPAPSRFEVFEKRETPKLI-I-NIKDYAINISOSEVYEEIGKIS 178
Qy      182 RILSKINOPYQKFLD--VLANTIKNAS--DSDGODLLETFNOLKE---HPTDFSVETLEON 234
Db      179 LDIIRK-----DKSIDPRELNLIKSISDDSDSLFSGKEFEKLELNKKSIDINFIKEN 233
Qy      235 SNEVGEFAKAFAYYIEPOHRDVLQLYAPEAFENYMDKENE 274
Db      234 LTFEQHAFSLAFSYFYAPDHRFTVLEIYAPDMFEYMKLEK 273

RESULT 3
STR_AQUAE
ID      SYR_AQUAE      STANDARD:      PRT:      583 AA.
AC      067068;
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      ARGINTL-TRNA SYNTHETASE (EC 6.1.1.19) (ARGININE--TRNA LIGASE) (ARGRS).
GN      ARGS OR AC_923.
OS      Aquifex aeolicus.
OC      Bacteria; Aquificales; Aquificaceae; Aquifex.
OX      NCBI_TaxID=63363;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-VF5;
RX      MEDLINE=98196666; PubMed=9537320;
RA      Deckert G., Warren P.V., Gaasterland T., Young M.G., Lenox A.L.,
RA      Gidman D.E., Overbeek R., Sneed M.J., Keller M., Aujay M., Huber R.,
RA      Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT      "The complete genome of the hyperthermophilic bacterium Aquifex
RT      aeolicus";
RL      Nature 392:353-358(1998).
CC      -I- CATALYTIC ACTIVITY: ATP + L-ARGININE + TRNA(ARG) = AMP +
CC      PHOSPHATE + L-ARGINYL-TRNA(ARG).
CC      -I- SUBUNIT: MONOMER (BY SIMILARITY).
CC      -I- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC      -I- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC      -----
CC      CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      CC      the European Bioinformatics Institute. There are no restrictions on its
CC      CC      use by non-profit institutions as long as its content is in no way
CC      CC      modified and this statement is not removed. Usage by and for commercial
CC      CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; AE000714; AAC07033.1; -
DR      InterPro; IPR001278; trna-synt_1d.
DR      InterPro; IPR001412; trna-synt_1.
DR      Pfam; PF00750; trna-synt_1d.1.
DR      PRINTS; PR01038; TRNASYNTHARG.
DR      PROSITE; PS00178; AA-TRNA-LIGASE.L; 1.
KW      Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW      Complete proteome.
FT      SITE      123 "HIGH" REGION.
FT      SITE      406 "KMSKS" REGION.
FT      BINDING    409 ATP (BY SIMILARITY).
SQ      SEQUENCE 583 AA; 68318 MW; 350807CA39C1220B CRC64;

Query Match      9.2%; Score 129.5; DB 1; Length 583;
Best Local Similarity 21.1%; Pred. No. 1.5;
Matches 68; Conservative 58; Mismatches 108; Indels 89; Gaps 12;
Qy      38 EKKKKKKKKRD---EERNKTOEHLKEIMKHVYKIE---VKGGEAVKKEAAKLE 89
Db      178 EKCPDEDETFEIKEIFEKDGVRGEYVKEIAERLKLIVGESICKPEANLKEVREKIIL 237

```

```

Qy      90 K-----VPSVLEMYKAIGKIIYVDDGIRKHSLEALSD-----KKIKD--- 131
Db      238 EESIELYTKKEIEPKKVDYDLISNYG--LDLMKEIKEDISLMDISFDWFESESLDSDGE 295
Qy      132 -----YGRDALL----- 149
Db      296 VERLILMLKEGYVYKDGALMLKTSFLPGDDKRVYKRSDDGTYYTPASDIAYHYNFKRG 355
Qy      150 YEPVLYIOSS--DYVENTEKALNYYEIGKIISRDILSKINOPYQKFLDY-----L 199
Db      356 FEKVIVMWGADHGGYIPRYKAALKM--LEIPEMLLEILLYQVMYLFREGKEVKKSKRAGTE 414
Qy      200 NTKNASDSDGOD--LTFNOLKEHPTDFSVETLEONSE-----VOEVFAKAFAYYTE 251
Db      415 VTIRELLDEYGDVAVFIFLTAKSDPDPDVEKAKESSENYVYOTAHARIISGIFRE 474

Qy      252 PQHRDVLQLYAPEAFENYMDKENE 274
Db      475 FKERYKKDVSVEELINYVQHLEE 497

RESULT 4
SMC1_YEAST
ID      SMC1_YEAST      STANDARD:      PRT:      1225 AA.
AC      P32908;
DT      01-OCT-1993 (Rel. 27, Last sequence update)
DT      01-OCT-1993 (Rel. 27, Last sequence update)
DT      15-JUL-1998 (Rel. 36, Last annotation update)
DE      CHROMOSOME SEGREGATION PROTEIN SMC1 (DA-BOX PROTEIN SMC1).
GN      SMC1 OR CHL10 OR YFL008W.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=5288C / AB972;
RX      MEDLINE=94103320; PubMed=8276886;
RA      Struninikov A.V., Lartionov V.L., Koshland D.;
RT      "SMC1: an essential yeast gene encoding a putative head-rod-tail
RT      protein is required for nuclear division and defines a new ubiquitous
RT      protein family.";
RL      J. Cell Biol. 123:1635-1648(1993).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=5288C / AB972;
RX      MEDLINE=95400292; PubMed=7670463;
RA      Murakami Y., Naitou M., Hagiwara H., Shidata T., Ozawa M.,
RA      Sasamura S.-I., Sasamura M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA      Yamazaki M., Tashiro H., Eki T.;
RT      "Analysis of the nucleotide sequence of chromosome VI from
RT      Saccharomyces cerevisiae";
RL      Nat. Genet. 10:261-268(1995).
CC      -I- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC      PART OF A CHROMOSOME CONDENSATION MOTOR.
CC      -I- SUBUNIT: HOMODIMER OR HETERODIMER WITH SMC2 OR OLIGOMERS.
CC      -I- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.
CC      -I- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC      FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC      -I- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC      -----
CC      CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      CC      the European Bioinformatics Institute. There are no restrictions on its
CC      CC      use by non-profit institutions as long as its content is in no way
CC      CC      modified and this statement is not removed. Usage by and for commercial
CC      CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; L00602; AAA16595.1; -
DR      EMBL; D50617; BAA09230.1; -
DR      EMBL; D31600; BAA06496.1; -
DR      PIR; A49464; A49464.

```

DR PIR: S41804, S41804.
DR SGD: S0001886; SMC1.
DR InterPro: IPR001687; ATP_GRP_A.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02463; SMC_C; 1.
DR Pfam: PF02463; SMC_N; 1.
KW Mitosis; ATP-binding; Coiled coil; Nuclear protein.
FT NP_BIND 33 40 ATP (POTENTIAL).
FT DOMAIN 173 489 COILED COIL (POTENTIAL).
FT DOMAIN 679 1063 COILED COIL (POTENTIAL).
FT DOMAIN 1057 1061 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 1137 1164 ALA/ASP-RICH (DA-BOX).
FT MOTAGEN 173 173 S->L: IN TS MUTANT SMC1-2.
FT MOTAGEN 458 458 N->D: IN TS MUTANT SMC1-1.
SQ SEQUENCE 1225 AA; 141279 MW; B504017AA0ECCAB8 CRC64;

Query Match
Best Local Similarity 8.7%; Score 122; DB 1; Length 1225;
Matches 62; Conservative 50; Mismatches 89; Indels 82; Gaps 13;

QY 32 VGMHVEKENKDNKDKERKKTQDEHLKEIKHIVKEVGE-EAVKKEAEKLLER 90
DB 796 IGTITEYENHSGELMKQ-----OSKEIQLOLOKILTEVKNLQFETDLSTTORRYEK 848
QY 91 VPSDVEEMVKAIGKIYVDGDTKHSLEALSEDKKKIKDIYDK-DALLHEHYVAKEG 149
DB 849 AQKD-LE-----NAQVEMKSLSEQEALEMKGSTESKEEKKHNLDEL 891
QY 150 YEPVVL-----IOSSEDEVENTERALNVYIEGKILSRDILSKINQYOKF-LDVLNTIKN 204
DB 892 QKKFVYKQSELSNSEDILSDMNSNL-----QVLKRE-RDGIKEIDKEFDLERRVALKN 943
QY 205 A-----SDSGQDLFTNQ-----KEHPTGFSVEFLQ 233
DB 944 CITSINIPISSETTIDLPISSTQNEATITSIDINYGKLPKRYKENNTSAREKLEQ 1003
QY 234 NSNEVOEVE-----AKAFAYIEPQHR-----DVLQIYKAE 264
DB 1004 KHEVEEILNELQPNKALERIDEAEGREYVINEETQLKAE 1046

RESULT 5
RBP1_PLAVB STANDARD: PRT: 2869 AA.
AC 000798:

DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
GN RBP1.
OS Plasmodium vivax (strain Belen).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9231538; PubMed=1617731.
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites";
RT Cell 69:1213-1226(1992).
RL -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -1- SUBUNIT: HOMODIMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.

CC EMBL: M88097; AAA29743.1; -
DR HSSP: P36956; 1AM9.
DR Malaria; Receptor; Signal; Transmembrane.
KW Malaria; Receptor; Signal; Transmembrane.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCF CRC64;

Query Match
Best Local Similarity 8.6%; Score 121; DB 1; Length 2869;
Matches 68; Conservative 42; Mismatches 128; Indels 78; Gaps 11;

QY 31 DVGHHVEKENKDNKDNK-----KDEERKKTQDEHLKEIKHIVK----- 71
DB 746 DINALIEVEKFEVTENKESTLEMLKDEE---MEERLQDAKETFAKLNFEVSDKLDIVYT 801
QY 72 ---EYKGEAVKKEAEKLLERKPSDVEEMVKAIGKI-----YVDDG-ITKHT 117
DB 802 KMSAEVTNAGIKKEIAQGFENVHKKMFSDAFSTKFPALONSQOITNOESDALEKHK 861
QY 118 SLEALSEDK---KTKIDYKDALLHEHYVAKEG---BPVLVIOSSSEDEVENTER 168
DB 862 QNRSEKEEVEFKNVESEDLSDRETEOEETKHKHNFRRKRGISAETNMREVLIKIS 921
QY 169 ALNVYIEI-----KTIISRLSLKINQYOKF-LDVLNTIKNA--- 205
DB 922 QLNIVYIEIKYPSLIDQNEVSTAKALKETIVSDSLRDIIDQYETFEKERTSAVENTVST 981
QY 206 --SDSGQDL-----FTNQLKEHPTGFSVEFLQNSNVOEVEFAKAFAYIEPQHRVL 258
DB 982 IQSLSKAIDSLKRLNGSINCKKYNTD--IDLLRSIKTLREYVQKEMKRGDKCGENT 1039
QY 259 QLYAPAEAFVMDKFN 274
DB 1040 ALLKSLRDKMGKINE 1055

RESULT 6
HS82_YEAST STANDARD: PRT: 709 AA.
AC P02829:

DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HEAT SHOCK PROTEIN HSP82.
GN HSP82 OR HSP90 OR YPL240C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84185722; PubMed=6325446;
RA Fairclly F.W., Finkelstein D.B.;
RT "Complete sequence of the heat shock-inducible HSP90 gene of
RT Saccharomyces cerevisiae";
RT J Biol. Chem. 259:5745-5751(1984).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RA Pohl T.M.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP ATPASE ACTIVITY.
RX MEDLINE=93123274; PubMed=8419347;
RA Nadeau K., Das A., Walsh C.T.;

```

RT 323 Hsp90 chaparrins possess ATPase activity and bind heat shock
transcription factors and peptidyl prolyl isomerases.";
RL J. Biol. Chem. 268:1479-1487(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE-97331324; PubMed-9187656;
RA Prodromou C., Roe S.M., Piper P.W., Pearl L.H.;
RT "A molecular clamp in the crystal structure of the N-terminal domain
of the yeast Hsp90 chaperone.";
RL Nat. Struct. Biol. 4:477-482(1997).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 2-214.
RX MEDLINE-97373820; PubMed-9230303;
RA Prodromou C., Roe S.M., O'Brien R., Ladbury J.E., Piper P.W.,
Pearl L.H.;
RT "Identification and structural characterization of the
ATP/ADP-binding site in the Hsp90 molecular chaperone.";
RL Cell 90:65-75(1997).
CC -1- FUNCTION: HSP82 IS AN ESSENTIAL PROTEIN THAT IS REQUIRED BY CELLS
IN HIGHER CONCENTRATIONS FOR GROWTH AT HIGHER TEMPERATURES.
CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY
(BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- INDUCTION: EXPRESSED CONSTITUTIVELY AT A VERY LOW LEVEL AND IS
STRONGLY INDUCED BY HEAT.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: K01387; AAA03743.1; -
DR EMBL: Z67751; CAA91604.1; -
DR EMBL: Z73596; CAA97961.1; -
DR PIR: A03313; HBBY90.
DR PDB: 1AH6; 22-OCT-97.
DR PDB: 1AH8; 22-OCT-97.
DR PDB: 1AM1; 24-JUN-98.
DR PDB: 1AMW; 24-JUN-98.
DR PDB: 1A4H; 05-AUG-98.
DR SWISS-2DPAGE: P02829; YEAST.
DR YEPD: 9884; -.
DR YEPD: 9908; -.
DR SGD: S0006161; HSP82.
DR InterPro: IPR003594; HATPase_c.
DR InterPro: IPR001404; HSP90.
DR Pfam: PF02518; HATPase_c; 1.
DR Pfam: PF00183; HSP90; 1.
DR PRINTS: PR00775; HEATSHOCK90.
DR SMART: SM00387; HATPase_c; 1.
DR PROSITE: PS00296; HSP90; 1.
DR Chaperone: ATP-binding; Heat shock; Multigene family; 3D-structure.
SQ SEQUENCE 709 AA; 81406 MW; D7C35676D68FB63 CRC64;

```

```

Query Match      8.58; Score 120; DB 1; Length 709;
Best Local Similarity 22.28; Pred. No. 6.3;
Matches 66; Conservative 48; Mismatches 91; Indels 92; Gaps 14;

```

```

QY 36 VKEKKNDKDEKRRKDEENKTOEHLKELMKHYIEVKGEEAEKAEKLEK----- 91
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 219 IPEEK-KDEK-KDE-KKDEDKK-----KLEVEDEEEKPKTKKKEVEGLE 268
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 92 -----PSDYLE-----MYKAIG-----KIYVGDGDTKHSLEA----- 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 269 ELNKTPLMTNPSDITQEEYNAPFKYSINDEWEDPLYV-----KHFSVEGLEFRALIF 322
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 -----LSEDKKIKIKYKGDALHHEHYAKGCEPVLV-----IQSSDYEN 165
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

DB 323 IPRAPDFLESKKKKNNI-----KLYRVFTTDEADLLPEMLSPKGVVDSFDPLN 377
QY 166 TEKALNYYVEIGKILSRDILSKI-----NQYOKFLDVLN-TIKNASDGDODLL 214
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 378 LSREHLQONKIMKIKRIKKNIIVKILAEFNEIAEDSQEFKFSAPFSKIKLGVHEDTORA 437
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 215 FTNQIKHEPTDSVEFLEQNSNEQVEFAKAFAYITFQHRDVQLYAPEAFNWDK 271
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 438 ALAKILLRYNSTKSVDELSTLDYVTRM-----PEHOKNIYITGESIKAVEK 484
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
REST_HUMAN STANDARD; PRT; 1427 AA.
ID REST_HUMAN
AC P30622;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RESTIN (CYTOPLASMIC LINKER PROTEIN-170 ALPHA-2) (CLIP-170) (REED-
DE STERNBERG INTERMEDIATE FILAMENT ASSOCIATED PROTEIN).
GN RSN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Peripheral blood monocytes;
RX MEDLINE-92289675; PubMed-1600942;
RA Bilbe G., Delabie J., Bruegen J., Richner H., Asselbergs F.A.M.,
RA Ceweloff N., Sorg C., Odink K., Tarsay L., Wiesendanger W.,
RA de Wolf-Peters C., Shipman R.;
RT "Restin: a novel intermediate filament-associated protein highly
expressed in the Reed-Sternberg cells of Hodgkin's disease.";
RL EMBO J. 11:2103-2113(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-92405160; PubMed-1356075;
RA Pierre P., Scheel J., Rickard J.E., Kreis T.E.;
RT "CLIP-170 links endocytic vesicles to microtubules.";
RL Cell 70:887-900(1992).
CC -1- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
CYTOSKELETON.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
OF HODGKIN'S DISEASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X64838; CAA46050.1; -
DR EMBL: M97501; AAA35693.1; -
DR PIR: S22695; S22695.
DR MIM: 179838; -.
DR InterPro: IPR000938; CAP-GLY.
DR InterPro: IPR001878; ZnF_CCHC.
DR Pfam: PF01302; CAP_GLY; 2.
DR SMART: SM00343; ZnF_C2HC; 1.
DR PROSITE: PS00845; CAP_GLY_1; 2.
DR Cytoskeleton; Microtubules; Coiled coil; Alternative splicing.
FT DOMAIN 60 125 CAP-GLY 1.
FT DOMAIN 143 204 SER-RICH.
FT DOMAIN 214 279 CAP-GLY 2.
FT DOMAIN 304 331 SER-RICH.

```

FT DOMAIN 350 1342 COILED COIL (POTENTIAL).
 FT DOMAIN 1408 1421 CCHC-BOX.
 FT VARSPLIC 457 491 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 1069 1069 D -> E (IN REF. 2).
 SQ SEQUENCE 1427 AA; 16098 MW; 0A4F166DD94254E8 CRC64;

Query Match

Best Local Similarity 8.4%; Score 118; DB 1; Length 1427;
 Matches 60; Conservative 51; Mismatches 99; Indels 66; Gaps 13;

QY 36 VKEKKNKDNKKR-----DEBNKQOEHLKIMKHYIKVKGGEAVK-----80
 DB 692 IREKENSLEAISKLDAEDHLEMDLNKQEAERK-VKELEVLAKNEQTKYID 749
 QY 81 -----KEAEKLE-----KVPDYLEMYKAIGGIIYVDGDTTKHISLEALSDKKK 128
 DB 750 NPTSQLKATEEKLDDALMRKASSSEKSMKRLRQOLEAEKQI-KHLEIEKNASSKAS 808
 QY 129 --IKIYKDALH-----EHYVAKEGYEPVLVIOSEDEVTEKALNYYEIGKITL 180
 DB 809 SITRELOGRELKTMLQENLSEVSQVKTLEKLOI-LKEKFAEASEBAVSQSRM-----863
 QY 181 SRDLISKINOPYOKFLDVLNTIKNASDSGDLLFTNQLKHPDTSVEFLQNSNEVOE 240
 DB 864 -QETVKNLHKEQEF-----NMLSSD-----LEKLRNLADMEKRFKEDEREQOL 908
 QY 241 VFAKAFAYVTEPOHVDLYLAPAFN--YMDKENE 274
 DB 909 IKAK-----EKLENDIAELMKSGDMSQTLTKMND 938

RESULT 8

RECN_AQUAE

ID RECN_AQUAE STANDARD; PRT; 520 AA.
 AC 066834;

DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE DNA REPAIR PROTEIN RECN (RECOMBINATION PROTEIN N).
 GN RECN OR AQ_561.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.
 RX NCBI_TaxID=63363;
 RN [1]
 RP SEQUENCE FROM N.A.

RA STRAIN-VES;
 RA MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anujay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.,
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 aeolicus.";
 RL Nature 392:353-358(1998).

-1- FUNCTION: MAY BE INVOLVED IN RECOMBINATIONAL REPAIR OF DAMAGED

DNA (BY SIMILARITY).

-1- SIMILARITY: BELONGS TO THE RECN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL, AE000695; AAC06789.1;
 DR InterPro: IPR003439; ABC_Transportr.
 DR Pfam: PF02483; SMC_C.1.
 DR DNA repair; ATP-binding; Complete proteome.
 FT NP_BIND 29 36 ATP (POTENTIAL).
 SQ SEQUENCE 520 AA; 60439 MW; A9708562ACBE901E CRC64;

Query Match
 Best Local Similarity 8.3%; Score 116; DB 1; Length 520;
 Matches 61; Conservative 52; Mismatches 78; Indels 48; Gaps 14;

QY 36 VKEKKNKDNKKRDEE-----RNKTQ-----EHLKIMKHYIKVKGGEAVKKAER 86
 DB 132 VKLEKVVNSLKRKEDELEFLRKKEELQKKLYLEFRAREVEIGISSEYEELNKKAN 191
 QY 87 L--LEKVPDYLE--MYKAIGK--IYVDGDTTKHIS-LEALSDKKRKIDYKDAL 139
 DB 192 LTNMLEKVKAKVGESEYIKLLEGENSEVYEIGIRKRLAVESYS--GKFSLEIKIAWL 248
 QY 140 HEHYVAKEGYEPVLVIOSEDEVTEKALNYYEIGKITLSRDLISKINOPYOK-FLDV 198
 DB 249 EE-----EYVE--LYNSLKEEMPELSEEVN--EINEKLF--TORLEEKYKSPPEI 295
 QY 199 LNTIKNASDSGDGDLFTNQLKHPDTSVEFLQNSNEVOEFAKAFAYVTEPOHND 256
 DB 296 LKEVE-----EIKEELSNLSVDFKKEELREVEKUREYKRLAEVSHD 340

RESULT 9

HS83_YEAST

ID HS83_YEAST STANDARD; PRT; 704 AA.
 AC P15106;

DT 01-APR-1990 (Rel. 14, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE HEAT SHOCK COGNATE PROTEIN HSC82.
 GN HSC82 OR YMR186W OR YMR010.16.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 RX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.

RA MEDLINE=89384620; PubMed=2674684;
 RA Borokovich K.A., Farrell F.W., Finkelstein D.B., Taulien J.,
 RA Lindquist S.;
 RT "hsp82 is an essential protein that is required in higher
 RT concentrations for growth of cells at higher temperatures.";
 RL Mol. Cell. Biol. 9:3919-3930(1989).

[2]

SEQUENCE FROM N.A.

STRAIN-S288C / AB972;

RA Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

[3]

SEQUENCE OF 1-7.

STRAIN-S288C;

RA MEDLINE=96093904; PubMed=7483834.

RA Boucherie H., Dujardin G., Kermorgant M., Monribot C., Slonimski P.P.,

RA Perrot M.;

"Two-dimensional protein map of Saccharomyces cerevisiae:

construction of a gene-protein index.";

Yeast 11:601-613(1995).

-1- FUNCTION: MOLECULAR CHAPERONE.

(BY SIMILARITY).

-1- INDUCTION: EXPRESSED CONSTITUTIVELY AT A VERY HIGH LEVEL AND IS

MODERATELY INDUCED BY HIGH TEMPERATURES.

-1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL, M26044; AA02813.1; -

DR EMBL: 249808; CAA89919.1; -
 DR PIR: A32572; A32572.
 DR HSSP: P02829; 1AH8.
 DR SWISS-2DPAGE: P15108; YEAST.
 DR YEPD; 9860; -.
 DR SCD; S0004798; HSC82.
 DR InterPro: IPR003594; HATPase_C.
 DR InterPro: IPR001404; HSP90.
 DR Pfam: PF02518; HATPase_C; 1.
 DR Pfam: PF00183; HSP90; 1.
 DR PRINTS: PR00775; HEATSHOCK90.
 DR SMART: SM00387; HATPase_C; 1.
 DR PROSITE: PS00298; HSP90; 1.
 DR Chaperone; ATP-binding; Heat shock; Multigene family; Repeat.
 FT INIT_MET 0
 FT DOMAIN 220 258
 FT 4 X 5 AA REPEATS OF [DE]-[DE]-E-K-K;
 FT HIGHLY CHARGED REGION.
 FT REPEAT 220 224
 FT REPEAT 225 229
 FT REPEAT 230 234
 FT REPEAT 245 249
 FT REPEAT 618 618
 FT CONFLICT 620 620
 FT L -> T (IN REF. 1).
 SQ SEQUENCE 704 AA; 80768 MW; 979B5258BDFD53B6 CRC64;

Query Match 8.3%; Score 116; DB 1; Length 704;
 Best Local Similarity 21.1%; Pred. No. 10;
 Matches 60; Conservative 52; Mismatches 101; Indels 72; Gaps 11;

QY 36 VKEKKNDENKRRKDEERK-----TOEHLKEIMKIHYIEKGEFAVYKEAEKLEKV 91
 DB 218 IPEEKKDEKDEKDEKDEKLEVEDEEKKPKTKKAAEEVDELELNK--TKPLMTN 275
 QY 92 PSDVLE---MYKAIIG---KIYIVDGDITKHISLEA-----LSKD 125
 DB 276 PSDITQEEYNAPFKSISNWMEDPLV-----KHFSVEQLFEPAFLIPKRAPDLPES 329
 QY 126 KKKIDYIGKDLLEHHYVYAKEGYEPVY-----IQSSEDIYENTKALNYEEIG 177
 DB 330 KKKKNNI-----KLYVRVFITDEAEDLIPWLSFYKGVYDSEDLPLNLSREMLQONKIM 384
 QY 178 KILSDILSKI-----NOPYOKFLDVL-NTIKNASDSGDGLFTPLNOKREPTDF 226
 DB 385 KYIRKNIYKLLLEAFNEIADSEDFKFSAFKNIKLVGHEDTONRALLAKILRNSTRK 444
 QY 227 SVEFLEQNSNEVOEYFAKAFAYIIEPQHRDVLQLYAPFAFNMYDK 271
 DB 445 SVDELTSITDVTYTRM-----PEHOKNIYITGESLKAVER 479

RESULT 10
 Y211_AQUAE STANDARD; PRT; 318 AA.
 AC 066405;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN AALL.
 GN AALL.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.
 OC NCBI_TaxID=63363;
 RN NCBI_TaxID=63363;
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Snead M.A., Keller M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 aeolicus.";

RL Nature 392:353-358(1998).
 CC -I- SIMILARITY: STROMG, TO A AEOLICUS AA07 AND AA34.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: AE000667; AAC07957.1; -
 DR Hypothetical protein; Plasmid; Complete proteome.
 KW SEQUENCE 318 AA; 38423 MW; 4C2E5B0050C6AC7C CRC64;
 SQ

Query Match 8.2%; Score 115.5; DB 1; Length 318;
 Best Local Similarity 25.8%; Pred. No. 4.3;
 Matches 59; Conservative 36; Mismatches 71; Indels 63; Gaps 14;

QY 32 VGMHYKEREKN-----KDEKKRDEERKTOEHLKEIMK-----IVK--IEYKGEAVK 80
 DB 53 VVHVHPVKKENLVLELDHRWKESKKREKELKELIIEYKNPDLIREILRLLEEGIR 112
 QY 81 KEAAEKLEKVPDVLNEMTKAIGKIYIVDGDITKHISLEALSDEKKIKDIYGDALDH 140
 DB 113 REKKDIAEKYKKEALELEERF--KPYLI-----KLRERLRKIN-----LQ 153
 QY 141 EHYVVA--KEGYEPVLIQSSDYENT-EKALNYYEIGKILSRDILSKINOPYOKFLD 197
 DB 154 ALYLLANVAE-----MFEKEEELKVMERAVKT-----ILFRQNKLOSP----- 195
 QY 198 VLNTIKNASDSGDGLFTPLNOLKHEPTDPSV-ELEQNSNEVOEYFAKA 245
 DB 196 -LGVLNK-----DFELP---KEPPYDFLSRFLQAELEPVLEKLKA 233

RESULT 11
 TPEF_CHICK STANDARD; PRT; 284 AA.
 AC P18441;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE TROPOMYOSIN ALPHA CHAIN, FIBROBLAST ISOFORM F1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN NCBI_TaxID=9031;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Body wall;
 RX MEDLINE=95120819; PubMed=7820856;
 RA Fanning A.S., Wolenski J.S., Mooseker M.S., Izant J.G.;
 RT "Differential regulation of skeletal muscle myosin-II and brush
 RT border myosin-I enzymology and mechanochemistry by bacterially
 RT produced tropomyosin isoforms.";
 RL Cell Motil. Cytoskeleton 29:29-45(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92084115; PubMed=1748294;
 RA Lemmonier M., Balvay L., Mouly V., Labri D., Fiszman M.Y.;
 RT "The chicken gene encoding the alpha isoform of tropomyosin of fast-
 RT twitch muscle fibers: organization, expression and identification of
 RT the major proteins synthesized.";
 RL Gene 107:229-240(1991).
 RN [3]
 RP SEQUENCE OF 189-213 FROM N.A.
 RC TISSUE=Muscle;
 RX MEDLINE=89345115; PubMed=2762137;
 RA Lemmonier M., Labri D., Fiszman M.Y.;
 RT "Chick alpha tropomyosin gene contains three sets of mutually

RT exclusive alternatively spliced exons."
 RL Nucleic Acids Res. 17:5400-5400(1989).
 CC -1- FUNCTION: THE FUNCTION OF TROPOMYOSIN IN SMOOTH MUSCLE AND NON-
 CC MUSCLE CELLS IS NOT CLEAR.
 CC -1- SUBUNIT: DIMER OF AN ALPHA AND A BETA CHAIN.
 CC -1- ALTERNATIVE PRODUCTS: THE MAJOR ISOFORMS OF SKELETAL MUSCLE,
 CC SMOOTH-MUSCLE, BRAIN AND FIBROBLAST TROPOMYOSINS ARE OBTAINED BY
 CC ALTERNATIVE MRNA SPLICING.
 CC -1- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
 CC EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
 CC -1- SIMILARITY: BELONGS TO THE TROPOMYOSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M36336; AAA65120.1; -;
 DR EMBL: X57991; CAA41058.1; -;
 DR EMBL: X57993; CAA41058.1; JOINED.
 DR EMBL: X57994; CAA41058.1; JOINED.
 DR EMBL: X57996; CAA41058.1; JOINED.
 DR PIR: S24401; S24401.
 DR HSSP: P03069; 1217.
 DR InterPro: IPR002017; Spectrin.
 DR InterPro: IPR000533; Tropomyosin.
 DR Pfam: PF00261; Tropomyosin.1.
 DR PRINTS: PR00194; TROPOMYOSIN.
 DR PROSITE: PS00326; TROPOMYOSIN.1.
 DR Coiled coil; Repeat; Alternative splicing.
 KW SEQUENCE 284 AA; 32960 MW; E94405DAB65597F CRC64;
 SQ
 Query Match 8.2%; Score 114.5; DB 1; Length 284;
 Best Local Similarity 22.9%; Pred. No. 4.3;
 Matches 52; Conservative 36; Mismatches 46; Indels 93; Gaps 9;
 QY 36 VKEKKNDENR-----KDEKNTQOEHLKETIMKIVIEVGEAVKKAEE 85
 DB 113 LEEAKKADESERGMKVTENRAQKDEKMEIOELKE-AKHI-----AEADAK- 161
 QY 86 KLEKVPSPDLEMYKAIGKIYVDGDTKHSLEALSDEKKIKIDYKDALHEHYV 145
 DB 162 -----YEEVARKLVITIGDLR-----AEERAE- 185
 QY 146 AKEGEPELVYQSSSEDEVTEKALNYYVEIGKIL--SRDILSKINOPYOKFLVLTNIK 203
 DB 186 -----SESQVROLEQLRINDQTLKALMAADKYSQKEDYEIEIKVL----- 228
 QY 204 NASDSGDGLFTNOLKEHPT-----DFSVFLEQNSMEVOEVFAKA 245
 DB 229 -----TDKLKAEETRAEFAERSVYKLEKSIDLDEKVAHA 263
 RESULT 12
 APM2_YEAST
 ID APM2_YEAST STANDARD: PRT; 605 AA.
 AC P38700;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE ADAPIN MEDIUM CHAIN HOMOLOG APM2.
 GN APM2 OR YHL019C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / REB526;

FX MEDLINE=95268148; PubMed=7749194;
 RA Stepp J.D., Pelliscena-Palle A., Hamilton S., Kirchhausen T.,
 RA Lemmon S.K.;
 RT "A late coiled sorting function for Saccharomyces cerevisiae Apm1p,
 RT but not for Apm2p, a second yeast clathrin AP medium chain-related
 RT protein."
 RL Mol. Biol. Cell 6:41-58(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=S288C / AB972;
 RX MEDLINE=94378003; PubMed=8091229;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
 RA Du Z., Favellio A., Fulton L., Gating S., Geisel C., Kirsten J.,
 RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
 RA Latreille P., Louis E.J., Macri C., Maridis E., Meneses S., Mouser L.,
 RA Nham M., Rifkin L., Riles L., St Peter H., Treviskis E., Vaughan K.,
 RA Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
 RA Vaudin M.;
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT VIII."
 RL Science 265:2077-2082(1994).
 CC -1- SIMILARITY: BELONGS TO THE ADAPTOR COMPLEXES MEDIUM SUBUNITS
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U09841; AAA83415.1; -;
 DR EMBL: U11582; CAB34896.1; -;
 DR PIR: S46833; S46833.
 DR SGD: S0001011; APM2.
 DR InterPro: IPR001392; Adap.comp.sub.
 DR Pfam: PF00928; Adap.comp.sub.1.
 DR PROSITE: PS00990; CLAT_ADAPTOR_M.1; 1.
 DR PROSITE: PS00991; CLAT_ADAPTOR_M.2; 1.
 KW Coated pits.
 SQ SEQUENCE 605 AA; 69990 MW; 7E216B11325EE3C CRC64;
 Query Match 8.0%; Score 113; DB 1; Length 605;
 Best Local Similarity 23.4%; Pred. No. 13;
 Matches 57; Conservative 31; Mismatches 72; Indels 84; Gaps 10;
 QY 37 KEKEKNKDNKRKDEERKKTQOEHLKEIYMKH-----IYKIEVGEAVKKAEE 85
 DB 175 KRDKRKKRKKRKKGKSGVSKLSIMVNNKKGINVEYKETLRNNDTGKEAAN 234
 QY 86 KLEKVPSPDLEMYKAIGKIYVDGDTKHSLEALSDEKKIKIDYKDALHEHYV 145
 DB 235 ---DELPNNDGLY-----INGDIKTI-----IMPISMRKGGH-----Y 267
 QY 146 AKEGEPELVYQSSSEDEVTEKALNYYVEIGKIL--SRDILSKINOPYOKFLVLTNIK 185
 DB 268 AKN--EFFLVYIERVYQTLMDERKGVIRKNLHGETVRCYLSGMPKLTISINKILNRDPQ 325
 QY 186 SKINOPYOK--FLDVLTNI-----KNASDS-----DGODLFTNOLKE 221
 DB 326 FMSNSSFHCQVSLDSINTIENDEKNSDDAGLQATDAREIEFIIPDGERVLCQVETLKR 385
 QY 222 HPTD 225
 DB 386 HVXD 389
 RESULT 13
 IF2_AQUAE
 ID IF2_AQUAE STANDARD: PRT; 805 AA.
 AC O67825;


```

QY 30 GDVGNHAKVEKE---KNKDNKRKDEERNKTOEHLKIMKHIVKTE-----VKGEPAVYKE 82
DQ 654 GEYRLLSIQHARKMSLOESMRERANKRTLEQDIDSLREBCAKKAHEHSAVNAEKKQ 713
QY 83 AAEKLEVPBVDLEWKAIGKTYIVDGTIKHIS-----LEALSEDKKIKIDYKDA 137
DQ 714 RAEELSRMFDSQMDLELREA-----HTROVSELRDLAKQHKHMDKDVHQKLL 762
QY 138 LLIHEHYAKGYEPVLVTOSSSEDEVTE-----KALN 171
DQ 763 LAHQQMT---ADYEVK-----ROEDAEKSSSELQNTILNREERQARKDLKGLDITVAKELQ 815
QY 172 VYVEIGKILSRDLKINQPYQKFLDVNTKNASDSGDGLLTNOLKEHPTFSVPEFL 231
DQ 816 TLHMLRKLFLVQDLOQRIRK-----NVVNESEEDGGSLAQKQ-----ISFL 857
QY 232 EONSNEVQEV 241
DQ 858 ENMLDOLTKV 867

```

```

RESULT 15
DYNAL_DROME STANDARD; PRT; 1265 AA.
ID DYNAL_DROME STANDARD; PRT; 1265 AA.
AC P13496; Q9YUAI;
DT 01-JAN-1990 (Rel. 13, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 150 KDA DYNEIN-ASSOCIATED POLYPEPTIDE (DP-150) (DAP-150) (GLUED
PROTEIN).
GN GL OR CG9206.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R, AND CANTON-S;
RX MEDLINE=87317680; PubMed=2819881;
RA Swarcop A., Swarcop M., Garen A.;
RT "Sequence analysis of the complete cDNA and encoded polypeptide for
the Glued gene of Drosophila melanogaster."
RL Proc. Natl. Acad. Sci. U.S.A. 84:6501-6505(1987).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wen K.H., Doyle J., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman J.P., Bhandari D., Bolintsov S.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Chu S.M., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertler S., Fleischmann W.,
RA Foster K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glisick A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin D., Houston K.A., Howard T.J., Hernandez J.R., Houck J.,
RA Jalali B.E., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasio P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,

```

```

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nussken D.R., Pacle J.M.,
RA Palazuelo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weisslock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000)
CC -!- FUNCTION: REQUIRED FOR THE CYTOPLASMIC DYNEIN-DRIVEN RETROGRADE
CC MOVEMENT OF VESICLES AND ORGANELLES ALONG MICROTUBULES. DYNEIN-
CC DYNACTIN INTERACTION IS A KEY COMPONENT OF THE MECHANISM OF AXONAL
CC TRANSPORT OF VESICLES AND ORGANELLES.
CC -!- SUBUNIT: LARGE MACROMOLECULAR COMPLEX OF AT LEAST 10 COMPONENTS.
CC P150(GLUED) BINDS DIRECTLY TO MICROTUBULES AND TO CYTOPLASMIC
CC DYNEIN.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.
CC -!- SIMILARITY: STRONG, TO OTHER SPECIES DYNEIN 150 KDA SUBUNIT.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC FRAMESHIFTS AT POSITIONS 32, 174 TO 220, 648 TO 672 AND 1208.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, J02932; -; NOT_ANNOTATED_CDS.
DR EMBL, AE003536; AAF49788.1; -.
DR PIR, A28313; A28313.
DR FlyBase: FBgn0001108; GI.
DR InterPro: IPR000938; CAP-Gly.
DR Pfam: PF01302; CAP-Gly; 1.
DR PROSITE: PS00845; CAP-Gly; 1.
KW Motor protein; Microtubules; Dynein; Coiled coil; Cytoskeleton.
FT DOMAIN 27 69
FT DOMAIN 105 138
FT DOMAIN 213 570
FT DOMAIN 812 836
FT DOMAIN 967 1084
FT DOMAIN 1128 1160
FT DOMAIN 1128 1160
FT CONFLICT 708 708
FT CONFLICT 875 875
FT CONFLICT 888 888
FT CONFLICT 1043 1043
FT SEQUENCE 1265 AA; 141217 MW; 2038A200282B2755 CRC64;

```

```

Query Match 8.08; Score 112; DB 1; Length 1265;
Best local Similarity 24.48; Pred. No. 35;
Matches 66; Conservative 48; Mismatches 111; Indels 46; Gaps 12;

```

```

QY 26 AGHGDMHAKVEKEKNKDNKRKDEE--RNKTOEHLK-EIMKHIVKIEKGEPAVYKE 82
DQ 370 SCGGDSGSLSTYEKQKOLEQONIRKLETLVRIDLSADHKDIOKLSKELEKRSEVELE 429
QY 83 -AAEKL-----LEKVPDLEWKAIGKTYIVDGTIKHISLEALSEDKKRT--KDIY 133
DQ 430 RTKEKLSAKIDLEAIVADLOEQVDALGAEWEQJAEKKMEL-----EDKVKLEEIEA 485
QY 134 GKALLHEHYVAKGYEPVLVTOSSSEDEVTEENTKALNYYE-----IGKILSRDLK 188
DQ 486 QLEALHEHYVAKGYEPVLVTOSSSEDEVTEENTKALNYYE-----IGKILSRDLK 188
QY 189 NOPYQKFLDVNTKNASDSGDGLLTNOLKEHPTFSVPEFL EONSNEVQEVFAKAFAY 248

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:48:11 ; Search time 112.1 Seconds

(without alignments)
514.086 Million cell updates/sec

Title: US-09-747-521-2_COPY_1_778

Perfect score: 3987

Sequence: 1 MNKKKEFIKIVSMSCLVTAI.....TSYGRNEAFPAFRLMH 778

Scoring table:

BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3987	100.0	809	22	WILD type B. anthr
2	3825	95.9	776	15	Lethal factor of B
3	1322	33.2	485	22	Lfn-Bcl-XL apoplos
4	1309	32.8	472	15	LF(1-254)-TR--PE
5	1307	32.8	508	15	LF(1-254)-TR--PE
6	1306	32.8	456	15	Adenyl cyclase gen
7	485	12.2	800	11	Plasmodium falcipa
8	253	6.3	1979	21	Plasmodium falcipa
9	228.5	5.7	1786	18	P. falciparum live
10	224.5	5.6	1558	21	Plasmodium falcipa
11	219.5	5.5	2663	22	Human polypeptide

12	219.5	5.5	2688	22	AAW40883	Human polypeptide
13	218	5.5	980	21	AAI18294	Plasmodium falcipa
14	215	5.4	1145	22	AAI82169	S. epidermidis ope
15	213	5.3	1516	21	AAI18195	Plasmodium falcipa
16	209.5	5.3	1780	22	AAI36881	Human polypeptide
17	202.5	5.1	1788	22	AAW40467	Human polypeptide
18	202	5.1	2710	17	AAI95016	C. difficile toxin
19	202	5.1	2710	17	AAI95016	Clostridium diffic
20	200	5.0	2485	21	AAI18172	Plasmodium falcipa
21	199.5	5.0	1254	11	AAI07503	Merzoite apical-en
22	199.5	5.0	1254	18	AAW24575	Human protein sequ
23	198.5	5.0	789	22	AAI95460	Human 160KD mediat
24	198.5	5.0	1427	12	AAI10534	B. burgdorferi ant
25	198	5.0	497	20	AAI20046	B. burgdorferi ant
26	196	4.9	481	20	AAI20047	Mouse RHAMM protei
27	195.5	4.9	630	18	AAW39166	Restin protein seq
28	195.5	4.9	1392	20	AAI06999	Kinetochore protei
29	194.5	4.9	3248	17	AAI99795	Maize RAD50. Zea
30	193	4.8	1316	22	AAI27248	Group B Streptococ
31	191.5	4.8	1164	17	AAI85781	Group B Streptococ
32	191.5	4.8	1164	19	AAW40537	Amino acid sequenc
33	191.5	4.8	1164	21	AAI84459	H. pylori cytoplas
34	191	4.8	2440	18	AAW20828	Mutant C-beta prot
35	190.5	4.8	1164	19	AAW40541	Rattus norvegicus
36	190	4.8	1886	16	AAW54241	Human RHAMM protei
37	189	4.7	725	18	AAI39165	Plasmodium falcipa
38	188.5	4.7	558	21	AAI18273	Mutant C-beta prot
39	188.5	4.7	1093	19	AAW40540	Human SCP-1 mutain
40	188	4.7	976	22	AAI66581	B. burgdorferi ant
41	188	4.7	1087	22	AAI19935	S. epidermidis ope
42	188	4.7	1119	20	AAI19934	Human polypeptide
43	188	4.7	5024	22	AAI82935	Mutant C-beta prot
44	187	4.7	1213	22	AAW40016	Human polypeptide
45	186.5	4.7	1099	19	AAW40538	Mutant C-beta prot

ALIGNMENTS

RESULT 1	
AAW47305	standard; Protein; 809 AA.
ID	AAW47305
AC	AAW47305;
DT	29-AUG-2001 (first entry)
DE	Wild type B. anthracis lethal factor.
XX	
KW	Lethal factor: LF; immunogen; LF4: protective antigen; PA: DNA vaccine;
KW	humoral; cell-mediated; immune memory response.
XX	
OS	Bacillus anthracis.
XX	
PH	Key
FT	Peptide
FT	Location/Qualifiers
FT	1..33
FT	/label= Signal peptide
FT	/note= "Not given in the specification"
FT	34..809
FT	/label= LF
FT	42..285
FT	/label= LF4
XX	
PN	WO200145639-A2.
PD	
XX	28-JUN-2001.
XX	
XX	21-DEC-2000; 2000WO-US34912.
XX	
PR	22-DEC-1999; 99US-0171459.
XX	
PA	(OHS) UNIT OHIO STATE RES FOUND.
PA	(GALL/) GALLOWAY D R.

PA (MATECZUN A J.
 XX Galloway DR, Mateczun AJ;
 XX WPI: 2001-408540/43.
 DR N-PSDB: AAC86015.
 XX
 PT Protecting animal against lethal infection with Bacillus anthracis, by
 PT administering wild type or mutated form of Bacillus anthracis lethal
 PT factor protein or its fragment or a nucleic acid encoding the mutated
 PT protein -
 XX
 PS Claim 3; Fig 1; 33pp; English.
 XX
 CC This sequence represents the B. anthracis lethal factor (LF). An
 CC immunogenic fragment of LF, LF4, can be used to produce an immune
 CC response which protects an animal against lethal infection with
 CC Bacillus anthracis. DNA encoding the B. anthracis LF can be used.
 CC In conjunction with DNA encoding the protective antigen (PA) in a
 CC DNA vaccine. Using a DNA vaccine which encodes the mutated LF protein
 CC or its fragment alone or in combination with a DNA encoding the PA protein
 CC or its fragment, both components (humoral and cell-mediated) of the
 CC immune system are stimulated, which results in longer term immune
 CC memory response. The combined use of a mutated LF and PA gene or their
 CC fragments results in a higher level of immune response, as judged by
 CC overall serum antibody titers for LF and PA antigens, than the use of
 CC either LF or PA genes in separate immunizations.
 CC
 SQ Sequence 809 AA;
 Query Match 100.0%; Score 3987; DB 22; Length 809;
 Best Local Similarity 100.0%; Pred. No. 1.1e-235;
 Matches 776; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 MNKKEFIKVSMSGLVATITLGSVPFIPLYOGAGHGDMHMKKKDKRDEE 60
 Db 1 mnikkefikvsmgclvatitlsgrvfiplyvgsgghgdmhmkkkdenkkrdee 60
 Oy 61 RNKQOEHLKIMKHIVKIEVKGEEAVKKEAEKLEKVPSPDLEMYKAGKITVDGD 120
 Db 61 rnkqoehlkaimkhivkievkggeavkkeaekllekvpdpvlemykagkityvvdgd 120
 Oy 121 ITRKISLASEDDKKTIDYGDALLHEHYVAKKEGEPPLVYQSSDYENNEKALNV 180
 Db 121 itrkislaseddkktidygkdallhehyvakegpeplvlyqssdyennekaln 180
 Oy 181 YYEIGKILSRDILSKINQYOKFLDVLNTIKNASDSDGQDLFTNQLKEHPTDSVEFL 240
 Db 181 yyeigkilsrdilskinqyokfldvlnitknasdsdgqdlftnqlkehpdtfsvefle 240
 Oy 241 QNSNEVOEVPKAFAYIIEPOHRDVLQLYAPAEFNYMDKFNEDTINLSLEELKQRMISR 300
 Db 241 qnsnevoevpkaafayiiiepohrdvlqllyapaeafnymdkfnedtinlsleelkqrmisr 300
 Oy 301 YEKKEKIKQOHQMSDSISEGKGLKKLQPIPEPKKDDIHSLSQEKELKLRQIDSS 360
 Db 301 yekkekikqohqmsdsiseegkglkkqlpipekkddihslsqeekekllkrtidss 360
 Oy 361 DFLSTEEKFEFKKLQIDIRDSLSEEEKELNRIQVDSNPLSEKEKEFLKRLKLDIQPYD 420
 Db 361 dflsteekfefkklqidirdslseeekekelnriqvdsnplsekekekllkrlkldiqpyd 420
 Oy 421 INORLQDTGLIDSPSINLAVRKQYKRDIONIDALLHQSISGTYLKNLYENNNINLT 480
 Db 421 inqrlqdtgglidspsinldvrkqykrdionidallhqsigsstlylknlyenninnlt 480
 Oy 481 ATLADLVDSDTNKINQINFEKKNKFKYSISSNMYTVDINERALLDNEKLKRIQISP 540
 Db 481 atlgadlvdsdtnkinqinfeekknkfkysissnmytvdineralldneklkriqlisp 540
 Oy 541 DTRAGYLENGKLILQNRNIGLEIKRQVQIKOSEKEYIRIDAKVVPKSIDTQIOEQLNIN 600
 Db 541 dtragylengklilqnrnigleikrqvqikosekeyiridakvvpksidtqioeqnl 600

Dg 541 dtragylengklilqnrnigleikrqvqikosekeyiridakvvpksidtqieeqnl 600
 Oy 601 QENNKALGLPKRYTKLIFENHNRYASINVESAYLILNEMKNNISDLIKRTNLYLVONG 660
 Db 601 qennkalgpkrytklifenvhnryasinvlesayyllnwmknnisdlikkrtnlylvong 660
 Oy 661 RFVFTDITLPNIAEQYTHQDEIIEQVNSKGLYVESRSILHLGSPKGYELANDSEGFHE 720
 Db 661 rfvtiditlpniaeqythqdeieqvnskgllyvesrsillhgspskgyelndsegfhe 720
 Oy 721 FGHAVDDVAGYLLDKNOSDLVTNSKFFDIFKEGNSNTSYGRTNEAEFFAERLHM 778
 Db 721 fghavddyagylldkngsdvltvnskkfddifkeegnsntsygtrneaeffaerlmh 778
 RESULT 2
 AAR60178
 AAR60178 standard; Protein: 776 AA.
 AAR60178;
 AAR60178;
 03-APR-1995 (first entry)
 Lethal factor of Bacillus anthracis.
 Anthrax; Bacillus anthracis; fusion protein; lethal factor;
 protective antigen; cell killing; targeting; targeting; pathogen;
 intracellular; HIV; human immunodeficiency virus; toxin.
 Bacillus anthracis.
 WO9418332-A.
 18-AUG-1994.
 14-FEB-1994; 94WO-US01624.
 12-FEB-1993; 93US-0021601.
 25-JUN-1993; 93US-0082849.
 (USSH) US DEPT HEALTH & HUMAN SERVICES.
 Atora N, Klmpel K, Leppla SH, Nichols PJ, Singh Y;
 WPI: 1994-279753/34.
 N-PSDB: AAQ70179.
 Nucleic acid encoding anthrax toxin fusion protein - useful for
 targeting toxin to specific cells, eg for killing tumour cells
 or HIV-infected cells
 Disclosure; Page 75-77; 124pp; English.
 The sequence encoding the lethal factor of Bacillus anthracis may be
 used in the construction of a nucleic acid which encodes a fusion
 protein comprising the anthrax protective antigen binding domain of
 the native anthrax lethal factor and a sequence encoding an activity
 inducing domain of a second protein. The fusion proteins are useful
 for the specific killing of tumour cells or the killing of cells
 infected with intracellular pathogens, especially HIV.
 Sequence 776 AA;
 Query Match 95.9%; Score 3825; DB 15; Length 776;
 Best Local Similarity 100.0%; Pred. No. 8.1e-226;
 Matches 745; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 34 AGGHDGMAHYKKEKKNKDEENKQOEHLKETIMKHIVKIEVKGEEAVKKEAE 93
 Db 1 agngdgmahykekekknkdeenkqoehlketimkhivkievkggeavkkeaee 93
 Oy 94 KLEKVPSPDLEMYKAGKITVDGDTITKHISLASEDDKKTIDYGRNALLHENHYV 153
 Db 94 klekvpdpvlemykagkityvddgtitkhislaseddkktidygknallhenhyv 153

|||||
Db kklekpsdvlemykaigklyivdgdtkhisllealsedkklkldygdallhenyvy 120
QY 154 AKGEYEPVLYIOSSSEDEVTEKALNVYVEIGKILSRDLISKINORYKRLDVLNTIKNA 213
Db 121 akegyepvlyiqssedvtekalnvyyeigkilsrdlisklnpqykildvlnlklkna 180
QY 214 SDSGDGLFTFTNOLKEHPTDFSVFELEONSNEVOEFAKAFAYIIEPOHEDVQLYAPEA 273
Db 181 sdsdgdglftftnglkehptdfsvfelegnsnegevafakafayiepqhndvqllyapea 240
QY 274 FNYMDFNEOEINLSLEELKDQRMLSRYEKMERIKOHYQWMSDSLSEEGRLKKLQIPI 333
Db 241 fnymdkfnegeinlsleelkdgqrmlysekwkikhqyhwdsdlsseegqllklkqip1 300
QY 334 EPKKDIIHLSQSEKELKRIODISSDFLSTKEFEFLKKLOIDIDISLSEEEKELNRI 393
Db 301 epkkdiihlsiseekellkriqidssdfsteekeflkkqididsiseekeellnri 360
QY 394 QVDSNPULSEKEFEFLKKLIDIOPIYDINOQLDGTGLIDSPSINLDRKQYKRDIONID 453
Db 361 qvdsnpulsekefeflkkikldioydingridgtgllidspsinldvrkqykrdionid 420
QY 454 ALHOSIGSTLTKIYLYENMNINLTATLGADLVSTDNFTKINRGIFNEFKKNFYSSIS 513
Db 421 allhosisgstlynkilyenmninltatlgadlvdstdnctknrgifnefkknfysis 480
QY 514 SNMVIYDINERPALDNERIKWRIQISPDPRAGYLENGKILLOINIGLEIKDVOITIOSEK 573
Db 481 snmviydinerpaldnerikwriqispdpragylengklllgrnigleikdvqllkqsek 540
QY 574 EYIRIDAKVPSKIDTKIOEAOINIOENWKALGLPKYTKLITFPVHNRYASINYESAY 633
Db 541 eyiridakvpskldtkqgeaglnindgenkalgpkkykllitfnvhnryasinyesay 600
QY 634 LLINEKNNIQSDLIKKTNYLVDSNGRFEVTDITLPNIAEQYTHODEIYEQVHSGLYV 693
Db 601 llineknniqsdlikkvtynyldvnggrfvtiditlpniaeqythdeleyeqvhsqlyv 660
QY 694 PRESRTLLGSPKGVLELRNDSSEGFHIEFGHAVDDYAGYLLDKQSDLYNMSKFFIDIFKE 753
Db 661 presrtllgspkgyvelrindsegfihiefgnavddyagylldknqsdlyvnskkffidifke 720
QY 754 EGSNLTSGRTNEAEPFAEAFRLMH 778
Db 721 egsnltsygrtneaeffaefrlmh 745

RESULT 3
AAU00222
ID AAU00222 standard: Protein: 485 AA.
XX
AC AAU00222:
XX
DT 31-MAY-2001 (first entry)
XX
DE LFn-Bcl-XL apoptosis-modifying fusion protein.
XX
KW Human; LFn-Bcl-XL; apoptosis; cancer; spinal muscular atrophy;
KW Anthrax lethal factor; neoplasm; tumour; hyper-proliferation;
KW Alzheimer's disease; neurodegenerative disorder; stroke;
KW transient ischaemic neuronal injury; spinal cord injury;
KW Huntington's disease.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Corynebacterium diptheriae.
OS Chimeric - Synthetic.
XX
FH Key
FT Location/Qualifiers
FT 5..10
FT /note= "6x histidine tag"
FT 21..276
FT /note= "Anthrax lethal factor amino acids 1 to 255"

FT Region 277..485
FT /note= "Bcl-XL amino acids 1 to 209"
XX
PN MO200112661-A2.
PD
XX 22-FEB-2001.
XX
PF 15-AUG-2000; 2000MO-US22293.
PR 16-AUG-1999; 99US-0149220.
XX
PA (HARD) HARVARD COLLEGE.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Youle RJ, Liu X, Collier RJ;
XX
DR WPI: 2001-218343/22.
DR N-PSDB; AAS00250.
XX
PT Novel fusion protein for modifying apoptosis in target cell and
PT reducing apoptosis after transient ischaemic neuronal injury, has two
PT domains which targets protein to a cell and modifies apoptotic response
PT of cell
XX
PS Claim 4: Page 64-65; 65pp: English.
XX
CC The sequence represents the amino acid sequence of LFn-Bcl-XL apoptosis-
CC modifying fusion protein comprising anthrax lethal factor (LF) sequence
CC fused to Bcl-XL. The functional apoptosis-modifying fusion protein is
CC capable of binding a target cell and integrating into or crossing a
CC cellular membrane of the target cell. The apoptosis-modifying fusion
CC protein comprises at least two domains: the DFR domain, which targets
CC the fusion protein to the target cell and the Bcl-XL domain, which
CC modifies an apoptotic response of the target cell. The fusion protein is
CC useful for modifying (inhibiting or enhancing) apoptosis in a target
CC cell, such as neuron, lymphocyte, cancer, neoplasm, macrophage,
CC epithelial, stem, tumour or hyper-proliferative cell or an adipocyte. It
CC is also useful for reducing apoptosis in a subject after transient
CC ischaemic neuronal injury, especially spinal cord injury. The fusion
CC protein may be used to treat various diseases and injury conditions
CC through inhibition or enhancement of apoptotic cellular response,
CC including neurodegenerative disorders such as Alzheimer's disease,
CC Huntington's disease, spinal muscular atrophy, stroke episodes and
CC unregulated cell growth as in tumours and various cancers. The apoptosis-
CC modifying fusion protein can be delivered effectively throughout the body
CC and targeted to selective tissue and cells.
XX
SQ Sequence 485 AA:
XX
Query Match 33.2%; Score 1322; DB 22; Length 485;
Best Local Similarity 58.8%; Pred. No. 4.3e-73;
Matches 293; Conservative 30; Mismatches 49; Indels 126; Gaps 11;
QY 34 AGGHGDMVHMKKEKKNKDEENKTOEHLKEIKHIVKIEVGEAVKKEAE 93
Db 22 agghgdmvhmkkekknkdeenktoehlkelnkhivklevgeavkkeaee 81
QY 94 KLEKVPDVLEMYKAIGKLYIVDGDITKHISLEALSSEDKKKIKIDYIGDALLHENYV 153
Db 82 kklekpsdvlemykaigklyivdgdltkhislealsedkklkldygdallhenyvy 141
QY 154 AKGEYEPVLYIOSSSEDEVTEKALNVYVEIGKILSRDLISKINORYKRLDVLNTIKNA 213
Db 142 akegyepvlyiqssedvtekalnvyyeigkilsrdlisklnpqykildvlnlklkna 201
QY 214 SDSGDGLFTFTNOLKEHPTDFSVFELEONSNEVOEFAKAFAYIIEPOHEDVQLYAPEA 273
Db 202 sdsdgdglftftnglkehptdfsvfelegnsnegevafakafayiepqhndvqllyapea 261
QY 274 FNYMDFNEOEINLSLEELKDQRMLSRYEKMERIKOHYQWMSDSLSEEGRLKKLQIPI 333
Db 262 fnymdkfnegeinlsleelkdgqrmlysekwkikhqyhwdsdlsseegqllklkqip1 275

Query Match	Best Local Similarity	Score	DB	Length
Matches 255; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	98.8%;	1309;	DB 15;	472;

Query Match	Best Local Similarity	Score	DB	Length
Matches 255; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	98.8%;	1309;	DB 15;	472;

07	34	AGGSGDVAMHYKKEKKNDEKKRDEENKPOEHLKIMKHIVKIKVGEAAVKKRAAE	93
D9	4	agghdyamhykekenkdenkirkdeernkqeehljelmhivk levbseeavkkaae	63
07	94	KLLEKVPSPVLEMYKATGKITIVDGDITKHISLEALSDEKSKKTKIDYGDALLHEHYV	153
D9	64	kllievpsdvlemykalygklyivdgdtklhlslealsedkklkldygdalldhehyv	123
07	154	AKEGEPVLVIOSSBDYVENETKALNVYETHGKILTSBIDLSIKINQPKFLDVNTKNA	213
D9	124	akegepvlvlyqssedyentekalnvyethgklltsldlsiklnpqkflvdvntlkna	183
07	214	SDSDGODLLFTNOLKEHPTDSVEFLEONSNEVOEVRKAKAFYITEPQHRDVLQLYAPEA	273
D9	184	sdsdgdqdllftnqlkehptdsvelfleonsnevevrfakafayitdpqhndvlyllyapea	243
07	274	FNMYMDKFNEOEINTLSLE 291	
D9	244	fnymdkfneoeintlstrae 261	

RESULT 5
 ID AAR60182 standard; Protein; 508 AA.
 AAR60182;
 04-APR-1995 (first entry)
 LF(1-254)--TR-PE(362-613) toxin fusion protein.
 Anthrax; Bacillus anthracis; fusion protein; lethal factor;
 protective antigen; cell killing; targeting; pathogen;
 intracellular; HIV; human immunodeficiency virus; toxin;
 Pseudomonas; exotoxin.
 Bacillus anthracis.
 Pseudomonas sp.
 WO9418332-A.
 18-AUG-1994.
 14-FEB-1994; 94WO-US01624.
 12-FEB-1993; 93US-0021601.
 25-JUN-1993; 93US-0082849.
 (USSH) US DEPT HEALTH & HUMAN SERVICES.
 Arcota N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;
 WPI: 1994-279753/34.
 N-PSDB; MA070183.
 Nucleic acid encoding anthrax toxin fusion protein - useful for
 targeting toxin to specific cells, eg for killing tumour cells
 or HIV-infected cells
 Example 1; Page 95-96; 124pp; English.
 This sequence is a fusion protein comprising amino acid residues
 1-254 of the anthrax protective antigen binding domain of the native
 anthrax lethal factor, a two residue linker and residues 362-613 of
 a Pseudomonas exotoxin A activity inducing domain of a second
 protein. Such fusion proteins may be useful for the specific
 killing of tumour cells or the killing of cells infected with
 intracellular pathogens, especially HIV, depending on their
 components.
 Sequence 508 AA;

Query Match 32.8%; Score 1307; DB 15; Length 508;
 Best Local Similarity 99.6%; Pred. No. 3.8e-72;
 Matches 254; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGGHGDVGMHVKEKKNKDEERNKTOEHLKEIMKHIVKIEVKGEEAVKKEAAE 93
 1 agghgdvgmhvkeknkdenkrkdeernktgeehlkeimkhivklevgeeeavkkeaee 60
 DB 1 agghgdvgmhvkeknkdenkrkdeernktgeehlkeimkhivklevgeeeavkkeaee 60
 QY 94 KLEEKPSDVLKEMKAIIGKITYVDGDTKHISLEALSEDKKIKIDYIGKDALLHEHYV 153
 61 kleeKpsdvlEmkai9gkIyIvdgdItKhIslealseDkkIkIdyIgdalIhehyv 120
 DB 61 kleeKpsdvlEmkai9gkIyIvdgdItKhIslealseDkkIkIdyIgdalIhehyv 120
 QY 154 AKEGYEPVLVIOSESEYVETTERKALNYYEIGKILSRDISKINOPYQKFLDVLNTIKNA 213
 121 akegyepvlviqssedyventekalnvyeyIgkIlsrdIsKInpYqkflDvlnItIkna 180
 DB 121 akegyepvlviqssedyventekalnvyeyIgkIlsrdIsKInpYqkflDvlnItIkna 180
 QY 214 SDSGDGDLFTNOLKEHPTDFSVFLEQNSNEVOEYFAKAFAYYIEPQHADVQLYAPEA 273
 181 sdsdgqdlIfTnqlkEhptdFsVfLeqnsnevQevfakafayyIepqhadvIqlyapea 240
 DB 181 sdsdgqdlIfTnqlkEhptdFsVfLeqnsnevQevfakafayyIepqhadvIqlyapea 240
 QY 274 FNYMDKFNQOEINLS 288
 241 fnymdkfnqeInlt 255
 DB 241 fnymdkfnqeInlt 255

RESULT 6

AAR60180
 ID AAR60180 standard; Protein; 456 AA.

AC AAR60180;

DT 04-APR-1995 (first entry)

DE LF(1-254)--TR--PE(401-602) toxin fusion protein.

XX Anthrax; Bacillus anthracis; fusion protein; lethal factor;

KW protective antigen; cell killing; targeting; targeting; pathogen;

KW Intracellular; HIV; human immunodeficiency virus; toxin;

XX Pseudomonas; exotoxin.

OS Bacillus anthracis.

XX Pseudomonas sp.

PN WO9418332-A.

PD 18-AUG-1994.

XX 14-FEB-1994; 94WO-US01624.

PR 12-FEB-1993; 93US-0021601.

PR 25-JUN-1993; 93US-0082849.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Aroza N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;

DR WPI: 1994-279753/34.

XX N-PSDB: AAR60180.

PT Nucleic acid encoding anthrax toxin fusion protein - useful for

PT targeting toxin to specific cells, eg for killing tumour cells

PT or HIV-infected cells

CC Claim 7; Page 86-87; 124pp; English.

CC This sequence is a fusion protein comprising amino acid residues

CC 1-254 of the anthrax protective antigen binding domain of the

CC native anthrax lethal factor, a two residue linker and residues

CC 401-602 of a Pseudomonas exotoxin A activity inducing domain of a

CC second protein. Such toxin fusion proteins may be useful for the

CC specific killing of tumour cells or the killing of cells infected

CC with intracellular pathogens, especially HIV, depending on their

CC components.

XX SQ Sequence 456 AA;

Query Match 32.8%; Score 1306; DB 15; Length 456;
 Best Local Similarity 100.0%; Pred. No. 3.8e-72;
 Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGGHGDVGMHVKEKKNKDEERNKTOEHLKEIMKHIVKIEVKGEEAVKKEAAE 93
 1 agghgdvgmhvkeknkdenkrkdeernktgeehlkeimkhivklevgeeeavkkeaee 60
 DB 1 agghgdvgmhvkeknkdenkrkdeernktgeehlkeimkhivklevgeeeavkkeaee 60
 QY 94 KLEEKPSDVLKEMKAIIGKITYVDGDTKHISLEALSEDKKIKIDYIGKDALLHEHYV 153
 61 kleeKpsdvlEmkai9gkIyIvdgdItKhIslealseDkkIkIdyIgdalIhehyv 120
 DB 61 kleeKpsdvlEmkai9gkIyIvdgdItKhIslealseDkkIkIdyIgdalIhehyv 120
 QY 154 AKEGYEPVLVIOSESEYVETTERKALNYYEIGKILSRDISKINOPYQKFLDVLNTIKNA 213
 121 akegyepvlviqssedyventekalnvyeyIgkIlsrdIsKInpYqkflDvlnItIkna 180
 DB 121 akegyepvlviqssedyventekalnvyeyIgkIlsrdIsKInpYqkflDvlnItIkna 180
 QY 214 SDSGDGDLFTNOLKEHPTDFSVFLEQNSNEVOEYFAKAFAYYIEPQHADVQLYAPEA 273
 181 sdsdgqdlIfTnqlkEhptdFsVfLeqnsnevQevfakafayyIepqhadvIqlyapea 240
 DB 181 sdsdgqdlIfTnqlkEhptdFsVfLeqnsnevQevfakafayyIepqhadvIqlyapea 240
 QY 274 FNYMDKFNQOEINLS 287
 241 fnymdkfnqeInlt 254
 DB 241 fnymdkfnqeInlt 254

RESULT 7

AAR04236
 ID AAR04236 standard; protein; 800 AA.

AC AAR04236;

DT 12-SEP-1989 (first entry)

DE Adenyl cyclase gene of Bacillus anthracis.

XX Adenyl cyclase; pertussis; protective vaccines; signal sequence

XX Bacillus anthracis.

OS EP366550-A.

PD 02-MAY-1990.

XX 25-OCT-1989; 89EP-0402949.

PR 25-OCT-1988; 88FR-0013952.

XX (INSP) INSTITUT PASTEUR.

PI Escuyer V, Duflot E, Mock M, Danchin A;

DR WPI: 1990-133988/18.

XX N-NSDB: 004123.

PT Nucleotide sequence encoding adenyl cyclase of Bacillus anthracis

PT and derived proteins, useful in protective vaccines, also effective

PT against pertussis

XX Claim 8; 13; 23pp; French.

CC In vivo the adenyl cyclase protein is synthesised as a precursor with a

CC signal sequence. The mature protein is secreted into the periplasmic

CC space, the signal peptide having been cleaved off at the moment of

CC secretion.

CC Given sequence contains several regions of close homology with the Cya

CC enzyme of Bordetella pertussis. Antibodies to B.anthraxis adenyl cyclase

CC cross-react with the enzyme from B.pertussis, hence a vaccine against the

CC former species will also protect against infection by the latter.


```

Db 469 ----lhjkeelke-svkltketqelgem-----vdikqeldqj-----gek 507
Oy 280 FNEOEINLSLEELKDOBMLSRY-----EKMEKIKOHYOMWSLSLEE----- 321
Db 508 ynaqlesstieslskekeynqkytleetlnlnekleetnkeytnlqnnynneimln 567
Oy 322 ----GRGLAKKIQPIEPKKDDI-----IHSLSQEEKEL---LKRQIDSSDPSTEER 368
Db 568 dlhmldngnktmktmctstlckndvhlneqldklnnekylnsksieelnvqindl--keek 625
Oy 369 EFLKLQIDIRLSLEEEKELLNRIQOVSSNPLSEKEKEFLKLKL----- 414
Db 626 dflnngivdlinsnqj-----dlittrkmeenkleenkykgemellrgnikssenlhn 680
Oy 415 DIOPVDINORL-----QDTGGLIDSPSILMDVRRKQKRIQINIDALLHOS 459
Db 681 deevcdlkrkkslkesemkmmkeehdkklaelkddcdv--lremeknedklnmlkee- 737
Oy 460 IGSTLYNKIYLENNMNNLNLTATLGADLVSDTDNTR-IRGIFNEFKKFKKYSISS--- 514
Db 738 -----yed-kin-----tlkeqnedkintlkeqnedkintlkeeyekhtntmkee 781
Oy 515 -NYMTVDINERPALD---NERLKMRIOLSPTFRAGYL-----ENGR 551
Db 782 yehkintlnegnehkintlnegnehkintmkeeyedkmtlnegnedkmslkeeyenki 841
Oy 552 LILQINIGLEIKDV---QIIKQSEKEIRIDAKVPRKSIDTNIQEQOLINQ----- 601
Db 842 nglnsneklkdvneyleevdklvtldex--kkqdkelnyahlkahneqillte 898
Oy 602 -EMNRALGIPKTYKLTITFVHNRYASNIYESAVYLILN-EMKNKIOS-DLIRKVTNVLVG 658
Db 899 meelkeqndkysdi-----yekyl-kikslcmintieccddeneditrrleeyinnn 952
Oy 659 NGRVFTDITLPLNIEQYTHODEIYEQVHSGKLYVPESRSILHGPSKVELRNDSEGT 718
Db 953 kgl-----kveeekelkr-----hsfnllkskef- 979
Oy 719 HEFGHVVDDYAGYLDKNOSDLYTNSKK 746
Db 980 --fknsleekshelkkhkekdllskdke 1005

```

```

XX 13-JUN-1995; .95FR-0007007.
PR (INSP ) INST PASTEUR.
PA Daubersies P, Druilhe P;
PI WPI: 1997-065464/06.
XX N-PSDB; AAT78868.
XX Plasmodium falciparum poly:peptide(s) and related nucleic acids -
PT derived from the liver stage antigen-3, useful for malaria vaccine
PR prodn. and diagnosis
XX
PS Claim 1; Fig 2A-I; 69pp; French.
CC This sequence corresponds to a Plasmodium falciparum strain K1
CC pre-erythrocytic liver stage antigen-3 (LSA-3) protein. The encoding
CC gene sequence was isolated by screening a P. falciparum strain 79/96
CC library with serum from a missionary treated by prophylaxis (for strain
CC T6/96 see FR9101286). Of 20 clones isolated, clone 729S was used to
CC screen a library generated from Thai strain K1. One clone contained a
CC 6.85 kb insert including the genomic sequence AAT78867. The gene
CC comprises a 1.8 kb region encoding 3 major blocks of tetrapeptide
CC repeats (especially the amino acid sequence VEEES, VEEEN, VEEI, VAPS,
CC etc) and a 3' hydrophobic region corresponding to a
CC glycosyl-phosphatidyl- inositol membrane anchoring sequence. The
CC invention relates to new polypeptides of at least 10 amino acids derived
CC from the LSA-3 protein with the exception of the peptides AAM24791-4.
CC The LSA-3 peptides can be used to raise antibodies and as vaccines for
CC immunotherapy of malaria.
XX
SQ Sequence 1786 AA:

```

Query Match 5.7%; Score 228.5; DB 18; Length 1786;
Best Local Similarity 18.9%; Pred. No. 1.5e-05;
Matches 155; Conservative 134; Mismatches 271; Indels 259; Gaps 32;

```

Oy 44 VKEKEKNKDENRKDEERKKTQDEHLKEIMKHVKLEVGEAVKKAELKEKPR--- 100
Db 891 leekleelhenvlsaalenqseekydvyl-----eevkeevatlletvegae 942
Oy 101 -----SDVLEMYKAIQGIY--IVDS-DITKHISLEALSE--- 132
Db 943 eksantitefleenaveseuvenaeenleknetvfnvldkveetveitsgeslennem 1002
Oy 133 DKKKINDIYKDALLHEHYVYAK-EGYEPVLVIQSS-----DVVENTEKAL 178
Db 1003 dkafseifdnvkgqenlltgmfrstetsivgsekvdenvenvsslldnlemkegl 1062
Oy 179 -----NRYVEIGKILSRD---ILSKINOPYOKFLDVLNVTIK 211
Db 1063 lnklenisstegvetelvehveqnyvvdvdpamkqfllglnaegglkemfnnledvfk 1122
Oy 212 NMSDSGQDLFLNOLKEHPTDPSVEFLQNSNEVOEFAKAFAYIIPQHRVLLVLP 271
Db 1123 ses-----dvilveelkdep-----vqkeveketvsilleemeenlvdlvee 1163
Oy 272 EAFNYMDKF-----NEOEINLSLEELKDOBMLSRYEKMEKIK 308
Db 1164 ekedldkmldaveesleisdsdskeeteslkdkvkslvveevqndmdesevkyleik 1223
Oy 309 QHTQHW-----SDSLSEEGRL-----LKLQIQPIE-----PKRD 338
Db 1224 nmeeelmkdaveindltsklleetgelneveadllkdmekilekalealsedskeildakd 1283
Oy 339 DITHSLSQEEKEL-----LKRQIDS-----SDFLSTEE-----KPF 370
Db 1284 dtlexkieehndlttldvevelkdvveedkiekvslkaleedllkevhekleeselle 1343
Oy 371 ----LKKLQIDIRDSLS-----EEKEELNRIQOVSSNPLSEKEKEFLKLKLDI 416

```


OY 652 TNYL 655
 1426 knfv 1429

RESULT 11

AAM39097 standard; Protein: 2663 AA.

AAM39097;

22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 2242.

Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokine; thrombolytic; drug screening; arthritis; inflammation; leukaemia.

Homo sapiens.

WO200153312-A1.

26-JUL-2001.

26-DEC-2000; 2000WO-US34263.

21-JAN-2000; 2000US-0488725.

25-APR-2000; 2000US-0552317.

09-JUL-2000; 2000US-0598042.

19-JUL-2000; 2000US-0620312.

03-AUG-2000; 2000US-0633450.

14-SEP-2000; 2000US-0662191.

19-OCT-2000; 2000US-0693036.

29-NOV-2000; 2000US-0727344.

WIPI: 2001-442253/47.

N-PSDB: AAI58253.

(HYSE-) HYSEO INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

Zhao QA, Zhou P, Goodrich R, Drmanac RT;

Novel nucleic acids and polypeptides, useful for treating disorders

such as central nervous system injuries -

Example 4; SEQ ID NO 2242; 10078bp; English.

The invention relates to human nucleic acids (AAI57798-AAI61369) and

the encoded polypeptides (AAM3642-AAM42213) with nootropic,

immunosuppressant and cytostatic activity. The polynucleotides are useful

in gene therapy. A composition containing a polypeptide or polynucleotide

of the invention may be used to treat diseases of the peripheral nervous

system, such as peripheral nervous injuries, peripheral neuropathy and

localised neuropathies and central nervous system diseases, such as

Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

utilisation of the activities such as: immune system suppression,

Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

and thrombolytic activity, cancer diagnosis and therapy, drug screening,

assays for receptor activity, arthritis and inflammation, leukaemias and

C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed

specification.

Sequence 2663 AA;

Query Match 5.5%; Score 219.5; DB 22; Length 2663;
 Best Local Similarity 20.3%; Pred. No. 9, 1e-05;
 Matches 168; Conservative 150; Mismatches 285; Indels 225; Gaps 39;

OY 42 MHVKEKKNK-----DENKRRDEBRNKTOE-EHLKEIMKH-----YKIEV 81
 Db 524 melkikexndldefealerktkkgemqllheisnlknkvkhyevngqdenelskvel 583
 OY 82 KGEAVKKEAAEKLLEKVPDVLEMYKAIIGKLIYIDGDTIKHLSLEASDEKKIK- 138
 Db 584 lre-----keqgkkkldqydsqklenlkm-----dis--ysleel edpkymkqtl 627
 OY 139 -----DIYGDALHEHYVYAKEGEPVLIQSESDYVENTERKALNYVEIGKILSR 190
 Db 628 fdaetvaldekresafirsenlelkemke--latykgmendldqygsleakkmy 684
 OY 191 DIISKINQPIQKFLDVINTKMASDSG---QDLFLFNQKHEPTDSVFE---LEQNSN 244
 Db 685 dlekelqsafnektltsll-----dgkvpkdlcnlelegktldqkelnkeveena 738
 OY 245 EVOEVF-----AKAFAYIEPQHHDV-----LDLYAPEAFNVMDF-----NEQELNLS 288
 Db 739 lreevillseikslpseverlrkelqkseehlhltse-----kklifsevvhvestrvqgl 794
 OY 289 LEEEL---KDQRLMSREKWEKIKOHYOMSDSLSEGRGLLKLQIPIEPKDDIHS-- 343
 Db 795 leelgtktkd-lattqsnysltqgefntf-----kclhndfeekymvleene 841
 OY 344 -LSQEREKLKRIQ-IDSS-DPLSTE-----EKEFLKLQIDIRD 380
 Db 842 rmngelvnlskaeqkfidsigalktelsyktqelqektrevegerlnemeqlke-qlenrd 900
 OY 381 S---LSEEEKELNRIOVDSNPLSEKEFEFLKKIKLIDQIPYDINQLOQT----- 428
 Db 901 splqtvereaktl-----itekltqtleevklitqekdkllqldgeslqterdqlk 949
 OY 429 GGLIDSPINLDVAKQYKRDQIONDALHOSIGSTLYLNKLYENMNINMLTATGADLV 488
 Db 950 sdldhvtvnmldtqgeqlrnaleslkq--hgetlntlksksevsrnlh-----me 998
 OY 489 DSTDNKIRNGIFNEFKKNKRYSSNYMTYVDINERPALDNERLKNRIQSPTRAGYLE 548
 Db 999 engetek-----defgk-----mvgldkkqgleakn----- 1025
 OY 549 NGKLLIQRNGLEIKDQIIRKQSEKREYIRIDAKVVPKSIDTKIQE-AQLINQENKAL 607
 Db 1026 -----tqltadvkdneileqqrkifslqeknelqgmlesvlaekekltokden--- 1076
 OY 608 GLPKYTKLITFNVHNRYSNIVESAYLLINEMKNNIQSDLIKVTNYLVPGNGRFVFTDI 667
 Db 1077 -----iemtleng-----eeltllgdelkk--qgelvaqekhnalikegelstcd 1120
 OY 668 TLPNIAEQYTHODEIYEQVHSKGLYVPEKRSILHGPSKQVELRNDSEGTIHEFGHAVDD 727
 Db 1121 rlaeveekllkexaqgkqgqllnvgeemseqkklneienlknkeltelhemete 1180
 OY 728 --YAGYILDKNQSDL--VTNSKKFI---DIFKEGSGNLSYGTGTNA 767
 Db 1181 rlelaqklnenyeovskitkerxvlkelqksfeterdhlygylelea 1228

RESULT 12

AAM40883 standard; Protein: 2688 AA.

AAM40883;

22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 5814.

Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX Homo sapiens.
XX OS
XX PN WO200153312-A1.
XX PD
XX PD 26-JUL-2001.
XX PF
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX
XX (HXSE-) HXSEQ INC.
XX PA
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX DR WPI: 2001-442253/47.
XX DR N-PSDB; AAI60039.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX PS
XX PS Example 2: SEQ ID NO 5614; 10078bp; English.
XX CC
XX CC The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AA163642-AA164213) with nootropic,
XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX CC
XX Sequence 2688 AA:
SQ

Query Match 5.5%; Score 219.5; DB 22; Length 2688;
Best Local Similarity 20.3%; Pred. No. 9.2e-05;
Matches 168; Conservative 150; Mismatches 285; Indels 225; Gaps 39;

QY 42 MHVKEKKNK-----DENKKRDEERNKTOE-EHLKEIMKH-----IVKIEV 81
DB 548 melkkekndidefealerktkkdqmqlhneisnlknlvkhreyngdlenelsskvel 607
QY 82 KGEAIVKKEAEKLEKVPDVLVEMKAIGKITYVDSDITKHISLEALSEDKKIK--- 138
DB 608 lre---kedqklkqeyidsqklenikm-----dl---ylesi-edbpkmqktl 651
QY 139 -----DIYGKDALLHEHYVYAKGEYEPVLVIOSSSEDYVENTEKALNVYIEGKILSR 190
DB 652 fdaetvaldakresafilsenlelkekme---latytkmendiqllysgleakkmqy 708
QY 191 DILSKINPYOKFLDVLTIRKNASDSG---QDLFTNOLKEHPTDFSEVF---LEONSN 244

DB 709 dekekgafneitkltstl-----dgkvpkdllicnlelegktdqlkelneveena 762
QY 245 EVOEVEF-----AKAFAYVIEPQHVDV-----LQLPAPAFNMDFK-----NEOEINLS 288
DB 763 lreevillselkslpseverlrkeigdkseehlitse-----kdlifevvhkeasrygl 818
QY 289 LEEL---KDQRLSRYEKWEKIKQHYQHWSDSLSEEGRLKLIQIPIEPKKDDIHS-- 343
DB 819 leeligktkdd-lattgsnykstidgefngf-----ktlhmdefqykmwleene 865
QY 344 -LSQFEKELKRIQ-IDSS-DELSFE-----EKEFLKLIQIDIRD 380
DB 866 rmngeivlnskeagkfdsllgalklelsyktqlqektreygerlnemeqlke-qleendr 924
QY 381 S---LSEEEKELNRIQVDSNPLSEKEKEFLKLIQIDOPIDINQRODT----- 428
DB 925 splqtverektl-----tektlqgtleevktlttegeddkqgdeslqdergdk 973
QY 429 GGLIDSPSINLDVRKQYRDIQNDIALHOSIGSTLYNKIYLYEMNNINMTATIGADLV 488
DB 974 sdhdtvnmldtqegqlnaleslky--hgetlnlkkakiseevsrnlh-----me 1022
QY 489 DSTDNTKINRGIFNEFKKNFKYSISNMIVDINERPALDNERLKWRIQLSPDTRAGYLE 548
DB 1023 entgetk-----defgqk-----awgldkkgldgaan----- 1049
QY 549 NGKLLIQRNIGLEIKDVOIKOSEKEYIRIDAKVVPKSKIDITQIE-AOLNINDEMNKAL 607
DB 1050 -----tqtltdvkdnelelqgrklfelqkneqlgmlesvlaekekltldken--- 1100
QY 608 GLPKTKLITFVNHKRYKSNVYESAVLLNKKNNIGSDLIKVTNYLYVDGNGRVPFDI 667
DB 1101 -----lemtleng-----eellrligdelk--qgeivageknhakkegelstrcd 1144
QY 668 TLPNIAEQYTHODEIYEQVHSGKLYVPESRSTILGPGKVELRNDSGFIHEGHAVDD 727
DB 1145 rlaeveeklkexgqlqegqqllnvgqeemgqklnelnlknelknelctlemete 1204
QY 728 --YAGYLLDKNQSDL--VTNSKKFI---DIFKEGSLNLTSGRTNEA 767
DB 1205 rlelaqklnyeeekstlkerkvklqeksfeterdhlryltelea 1252

RESULT 13
AAB18294
ID AAB18294 standard; Protein; 980 AA.
XX
XX AAB18294;
XX AC
XX 07-NOV-2000 (first entry)
XX DT
XX XX
XX DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:152.
XX XX
XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX KW antimalarial; malaria; protozoacide; infection; insecticide.
XX OS
XX OS Plasmodium falciparum.
XX PN
XX PN WO200025728-A2.
XX PD
XX PD 11-MAY-2000.
XX PF
XX PF 05-NOV-1999; 99WO-US26796.
XX PR
XX PR 05-NOV-1998; 98US-0107131.
XX PA
XX PA (HOEF/) HOFFMAN S.
XX PA (GARDU/) GARDNER J C.
XX PA (GARDU/) GARDNER M.
XX PA (VENT/) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;

XX WPI: 2000-365347/31.
DR Proteins encoded by chromosome 2 of the human malarial parasite,
XX Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P. falciparum infection.
PS Disclosure: Page 362-365; 577pp: English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic life cycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAH70078 to AAH70287 and AAH18144 to AAH18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX Sequence 980 AA:

Query Match 5.5%; Score 218; DB 21; Length 980;
Best Local Similarity 21.2%; Pred. No. 3.2e-05;
Matches 150; Conservative 126; Mismatches 262; Indels 168; Gaps 32;

OY 51 KQENKRRDEBRRKTOEBEIKKHIVKIEVGGEAVKREAAKLEKVPDYLEMYKAI 110
DB 345 KQENKRRDEBRRKTOEBEIKKHIVKIEVGGEAVKREAAKLEKVPDYLEMYKAI 110
OY 111 GGRYIVDPDITKHSLEALSEBKKIKDIYQKDALHHEHYVAKGEYFVLVIOSEEDY 170
DB 396 -----ekkkienl-----kdelin-----ikkmedkmmtemndl 427
OY 171 VENTKALVYVEIGKISRLDLSKINOPYQKFLDVLTIKNASDSGDGLFTNOLKEH 230
DB 428 lankveeln-----rlnkyeknlveln-----neldivlkkklndeeflke- 468
OY 231 PPDFSVFLEQNSNEQVEFAKAFAYIERQHDVQLYAPEFNMDFNEDEINLSL- 289
DB 469 -----eeekknldmwy-kikeyeiqike-----neidslkkneqnlhvl 508
OY 290 ---BELKIDQPMIL--SRYEKWEIKQHOHWSDSLSEGRGLKLRQPIPEPKDIIHLSL 345
DB 509 kneelnekeiilknydk--elnmleqynkkikeekdmnnkiksmgdtlhqng-leeemg 555
OY 346 QDEKELLRKIQ-----IDSSDFLSTEEREFKLKQIDIRDSLSEBEKELLNRQ-VDSS 398
DB 566 eenkkelkrlklnvcdmnlsgqllikenekhmqekve-eyknlkqgdqeklnliqydr 624
OY 399 NLFSEKEXEFL-----KKLKLDQRPYDINQRODQNGGLIDSPSINIDVRK 443
DB 625 lqignemedyndceekikgakinnkklitcatmanannmmndenlkekdkkin-dlmk 663
OY 444 QYKRDIONIDLAL-----HOSIGSTLYNKIYVENMNIINLTATIGADVLSTONTK 496
DB 684 dnekkkeeklnlveeksklehshv--kignemslvegn-eklkeemgslrlkimeel 740
OY 497 NNGI--FNEFKKNFKYSISNTYIVDINERPALDNERLKWRIQLSPDRAGYLENGKLI 554
DB 741 kdmekyeekkkneeerkkneeerkkneeerkkneeeekkkne--eerkkneeeekkklek 798

OY 555 QNRNGLEIKD-VOLIKOSEKEIYRIQAKVYPKS-----KIDPKIOEAQNLINQEW- 603
DB 799 dknhfeekermeyekedrkdkdkkxghssdkkeekynkkekckekssnllfdeeyl 858
OY 604 -----NRALGDPKTKLTITFENVHNRVYS-----NYESAVLIINEN 639
DB 859 lqleelrdtgencfiyikskiskeldv-lnkllkskdallndafnknlatstwnlfn- 916
OY 640 KNIQSDLIKRVNTNYLVDSNGRFEVFTDYL-PNIAEQYTHODEIYE 684
DB 917 -enkegnlftventategn-----ltdentevemneeyk 954

RESULT 14

AAH82169
ID AAG82169 standard; Protein; 1145 AA.

XX
AC AAG82169;

XX
DT 03-SEP-2001 (first entry)

XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:1432.

XX
XX Staphylococcus epidermidis SRI strain; Infection; diagnosis;
KW vaccination; endocarditis.

XX
OS Staphylococcus epidermidis.

XX
PN W0200134809-A2.

XX
PD 17-MAY-2001.

XX
PE 09-NOV-2000; 2000WO-US30782.

XX
PR 09-NOV-1999; 99US-0164258.

XX
PA (GLAXO) GLAXO GROUP LTD.

XX
PI kimerly WT;

XX
DR WPI: 2001-316495/33.

XX
DR N-PSDB; AAH53019.

XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -

XX
XX Claim 18: Page 405; 2188pp; English.

CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce host cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.

XX
XX Sequence 1145 AA:

```

Query Match      5.4%; Score 215; DB 22; Length 1145;
Best Local Similarity 19.2%; Pred. No. 5.9e-05;
Matches 161; Conservative 154; Mismatches 271; Indels 254; Gaps 38;

QY 41 GMHVKEKKNKDEKRRDEKRNKTOE-----EHKIKIMHVIKIEVGEAKKKEAE 93
    | | | | | : : : | : : | : : | : : | : : | : : | : : | : : |
Db 361 gffnkrkkkkeeqlkrlneatltqekvngvprplredmsfltrflok--havnekl1n 418
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 94 KLEKVPDVLEMY-----KALGKIYIVDGLT-----KHISL 127
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 419 qeydvvpelllselyqtqtslntyltfsdevlkalnkk1---enesplfeavnhqv 475
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 128 FALSEDKKKIDYIG-----KDAL-----LHEHYVAKGEPVLTQSSDEVNTE 175
    | | | | | : : : | : : | : : | : : | : : | : : | : : | : : |
Db 476 nelsdededyeydryelntlkslshnkyhlyhldslsk--ligrteftelkq 533
    | | | | | : : : | : : | : : | : : | : : | : : | : : | : : |
QY 176 KALNVYEIGILSRDISKINOPYOKFLDVL-----NRIKNASDS-----DGO----- 219
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 534 enstcaynhkhetqhnfevtstngdkraldivkvpilfrtkqdtldtllldnqtkvg 593
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 220 -----DLFTNOLKEHPPTDFSVFLEQNSNEVOEYAKAFAYIEPQHRDV 265
    | | | | | : : : | : : | : : | : : | : : | : : | : : | : : |
Db 594 vfgtfsagksallnallgenylvspsptaatetelsygsqtltkesqllle-eynhv 652
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 266 LQLYAPEAFNMKFNQGEIN--LSLEELKQD--RMLSRKEKEKIKOHYQHMSSDSISEGR 323
    | | | | | : : : | : : | : : | : : | : : | : : | : : | : : |
Db 653 lrfy-elsfnldfdesdldkllkleqnlafisalek-----hyemysml----- 700
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 324 GLKKLQIPPKKDDIHSISOEKEELKRIQIDSSDFLSTEKEEFLKQID----- 377
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 701 -----ehslhtvsleeik-----kwsadeyaltfktvnlklpdlaklki 742
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 378 IRDSL-----SEEEKELNRIQVSSNPL-----SEKEKEFLKKLKDIOPYDINQ 423
    | | | | | : : : | : : | : : | : : | : : | : : | : : | : : |
Db 743 lldslglsnmgqrhtneeqlltsedllytyfnhsftndkafiehm-----dmng 796
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 424 RLQOTGGLIDSPSINLDVRKRYKRDIONIDALHNSIGS--TLNKKITLY-----ENM 474
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 797 lneqatfmlinaavl--aedkqdlqavedyvadalqgvnlhseilyvsarsqlngnl 853
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 475 NINMLTAT-----NTRKNGRITNEFKKFKFSISSNMYIWDINE 523
    | | | | | : : : | : : | : : | : : | : : | : : | : : | : : |
Db 854 glnlreslsgfakvearltlegqmtylqgmntsfkmlkdfndnakklaarqklhny 913
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 483 -----LGADLVDSFD-----NTRKNGRITNEFKKFKFSISSNMYIWDINE 523
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 914 knqrtrlngealldtaqrtnfeveeqvylhnerlkqlldexkvsfnsqmtqnn--dfne 970
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 524 RPAIDNERLKRIOLESPTRAGYLENGKLIQOR-----NIGLETKDYOIITOSEKEYIRI 578
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 971 -----ekkiskiyldqihqrlfleg--slterlkkkyfnsqleeqilpvmkklnglhwil 1024
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 579 DAKV-VPKSKIDTKIOEQALNINQEMKALGPXY--TKLITFNVRHNRYSNIVESAYL 634
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 1025 naktfnvpsvtdrpllqlqeln-----smqlgslpqlktrkikvnpksqcdiqehangle 1079
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 635 ILNEMKNNIQ--SDLIKRVTVNLVDNGRFEVFTDILPNIAE--QYTHODEIYEQVHSK 689
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 1080 llqgdlnsltrqlndylnemtcg--laebqfqlmetsiq--qgidelstfiddtlllqqlakl 1137
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |

```

```

XX XX Plasmodium falciparum.
OS OS
XX XX
PN PN W0200025728-A2.
XX XX
XX XX 11-MAY-2000.
XX XX
PF PF 05-NOV-1999; 99WO-US26796.
XX XX
PR PR 05-NOV-1998; 98US-0107131.
XX XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX XX
DR DR WPI: 2000-365347/31.
XX XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX XX
PS Disclosure: Page 120-124; 577pp; English.
XX XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAAT0078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX XX
SQ Sequence 1516 AA:

```

```

Query Match      5.3%; Score 213; DB 21; Length 1516;
Best Local Similarity 19.8%; Pred. No. 0.00011;
Matches 162; Conservative 136; Mismatches 271; Indels 248; Gaps 37;

QY 45 KEKEKN-----KDEKNKDEERNKTOEHLKEIM-----KHIVIEVKEGEAVAKE 90
    | | | | | : : : | : : | : : | : : | : : | : : | : : | : : |
Db 677 kdlskntnlnlkdynvlgkkskkkkkflndlnlytnfteskygqllyvkgees--ke 734
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 91 AAEKLEKLV-----PSDVLMEYKAGIKIYIVODDIRKHISLEALSDX 134
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 735 dlknqldfvtqecyrnmdltdnksdflknklidmkkyl---ynlejeqeelnekk 791
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 135 K-----KINDYGDALLHEHYVAKEGEYPLVLTQSSDEVNTEKALNV 180
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 792 nynknmdsnkftflkietekalllddsqfsgslldlkeynytadndnnems1 851
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 181 YVEIGILSRDISKINOPYOKFLDVLFTIKNASDS--DCODLFTNOLKEHP--TDFS 235
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 852 yedgenfltr-----neplcneyeeknlllylsdeqkyneedlfxkklkekeknnds 905
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 236 VEFLEQNSNEVOEYAKAFAYIEPQHRDVLQVLAPEAFNMKFNQGEINUS-----LEE 291
    : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |

```

```

Db 906 saddfence--vqe-----kiyvnekieynknndksssssiilee 945
QY 292 LKDOR-----MLSNYEKWEKIKOHYQHWMS-----DSLSEGRGLKK 328
Db 946 lkykkekdelvspnlcvlldefehsndlenyisvssddmktvskniltgvkenkydk 1005
QY 329 LQIPIEPKKDDIHHSLSOEKEKELLRIOIDSSDFLSTEEKEFLKKLQIDIRDSLSEERE 388
Db 1006 tnveydkkgddgvleisfedshkleeskfdannilynddeleknl--kdyisdvdkn 1062
QY 389 LLMRIQVDSNPLSEKEKEFLKKLKDIOPYDINQRLQDTGGLDPSI--NLDVRQY 445
Db 1063 hvnnl-yntergederenef-----venkigtseehksenefictenkslrkqy 1109
QY 446 --KRDIONIDLHQSISTGLNKLTYLXENMNINLTLATLGAD-LVDSTDNTKINRGIFN 502
Db 1110 mskeelshnvrl-----ksddlnlksqnyfeilldk-----k 1142
QY 503 EFKKNFKYSISSNWMIVDINERPALDNRLKWRIOQSPTRAGYLENGKLIQRNIGLEI 562
Db 1143 qymdnfmgmlegn-----ndklyk-edkldegayfeylednkil----- 1179
QY 563 KDVOIIRKQSEKEEYIRIDAKVVPKSKIDYRKIOEAQULINQEWNKAL-----GLPKYTKL 615
Db 1180 --dsyiketnke---neelikeyk---klkknleindemnddikllnffgipyigsp 1230
QY 616 ITFNVHNRYASN-----IVESAYLLINEMWKNINQSDLIKRVNTNYLYDGNRGREYFTDIT 668
Db 1231 ceaeaqcsylhnhkycdaaisddsdvlfvsgkvtlknfknkkt----- 1274
QY 669 LPNIAEQYTHODEIYEO--VHSK-GLVVPESRSI-LHGSPS-----KGVELRNDSEGFTH 719
Db 1275 -----veyekakeekiglygeelnlisllcgcdytlvgvhgigivnale-lik 1322
QY 720 EFGHAVDDYAGYLLDKNOSDLVTNSKRFID--IFKEE 754
Db 1323 afpntfed-----lklldivsnpfkrldkmynee 1352

```

Search completed: December 2, 2001, 13:48:20
 Job time: 140 sec

THIS PAGE BLANK (USPTO)


```
Db 1111 EMSGCKKKEKKKKKISLINYEKKLONIIIDENKLENCITTEKKLAKMEDITKYINEQ 1170
QY 141 YGKDAL-----LHEHYVAREGEYPLVIOSSSEDEVENTERKALNVEYIGKILSRDILSK 195
Db 1171 FSLSKIOFENKMMNEVIFLKKKDESIYML---KELIKEKETILYNONILKQYKKDDDI 1227
QY 196 INOPYOKFLVYLVNTIKN-----ASDSOGDILLFTNOLKEHPDSEVSEVLEQNSNEVOEYFA 251
Db 1228 LKEMIEKIDIKKKLQOEIIISQKDRQIETLENLK-----IGKEKINKDNENIQRKQY 1282
QY 252 KAFAYIEPQ-----HRDVLQVAPFAFNYMDKFNQEOEINL--SLEE-LKDRMLSK----- 299
Db 1283 K-INIHLEKEIEKKTNDIEKEKTKLSNKFIDINKENNNLHMKIEFLKLEKEISTENV 1341
QY 300 -----RYEKMEKIKQHYOHMSDSLSEBGRLLKQIPIEPK 337
Db 1342 KLIETNKLYIENEKLSNDLKNLTKLEKEKIOENY-----NKINKHIKITIELQSYKEIKE 1397
QY 338 DDI-----IHSLSQEKELLRQIOWSDPLSTEEKEFLKLQIDIRDS--LSSEE 386
Db 1398 KHKMEITQIKQOIHNLDKH---IAQIQIEKNF---EESYLEKKNENKMSMILEEKY 1449
QY 387 KEL-----LNRIOY-----DSSNPLSEKEKEFLKKLQIDQPIORLOOTDGLID 433
Db 1450 KELSTYEDKNINKIKIEDEKKNENILTKNEE-INNLKEEYKM--VOQHLEDITNVLYE 1506
QY 434 SPSSINDLVRKQYKRDIONIDALLHOSIGSTLYNKIYLYENMN-----INNLFTATIGADL 487
Db 1507 KQKLAIDITKEKNNIINECDKIKNN--KLLNNK--LKEQNNYEHTLNNI----- 1554
QY 488 VDSIDNFKIRGIFNEKKNFKYISS-----NY-MIYDINRPALDERLKWRIOLS 539
Db 1555 -----KKEOQIIEERKKETOKVESLEHAFKOSYNOLKQONENLOOQIKQL--ANQOD 1607
QY 540 PDTRAGYLENGKLLIQNIGLEIKDVOIIEKSEKEYIRIDAKVYKPKIDITKIOEAOILI 599
Db 1608 IKTKSKMLKNNNEL-----IKETKNYSQOKKEFKIK-GLKNKQAVIKILKNENQOLKI 1659
QY 600 NO-EMNKALGLPKYTKLITFNHNRVANSIVESAYLILNEMKN-NIGSDILKQVNTYLV 657
Db 1660 NAEFYIKRDVODNY---VTLNVHNN-----ILNEQKFLFQIDILKS---QVD 1701
QY 658 GNGFVFTDITLPIAIOYHODEIYQVHSGLYVPESRILHSGSKGELRN----- 712
Db 1702 QKONIT-----NNMKQO-----IEDVNHKIASINKEEL---NTTKIKNKITED 1744
QY 713 ---DSEGFIEHFAVDYAGYLLDKNQSDLVTSKSKFIDIFKEGSNL--TSYGRTEAE 768
Db 1745 VNLSEVKLSLNSKDEVKKRTIEIKQKE--REYKKLLDQYIEKKNLVTKYK--ELD 1800
QY 769 FFAEAFRLMHS--TDHAERLKVNQNAKPTFOFINDOI 803
Db 1801 SYMTKYEFHAHAKYQYEEIEKDKNLK---LKDEV 1833

RESULT 2
Q9BJYO PRELIMINARY; PRT; 2752 AA.
AC Q9BJYO;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE 235 KDA RHOPTRY PROTEIN (FRAGMENT).
GN IIA.1.
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YM;
RA Khan S.M., Jatta W., Peter P.R.;
RT Distribution and Characterization of the 235 kDa Rhoptry Multigene
RT Family within the Genomes of Virulent and Avirulent Lines of
```

```
RT Plasmodium yoelii y.;
RL Mol. Biochem. Parasitol. 0:0-0(2001).
DR EMBL; AF323442; AAK15625.1; -.
FT NON_TER 1
FT NON_TER 2752
SQ SEQUENCE 2752 AA; 322606 MW; E834E3753168AF76 CAC64;

Query Match 6.6%; Score 273.5; DB 5; Length 2752;
Best Local Similarity 19.6%; Pred. No. 0.0018;
Matches 191; Conservative 158; Mismatches 283; Indels 343; Gaps 44;

QY 58 DEENKQOEELKIMKHYIKVEKGEBV-----KKEAAEKLKVSQVLEMKAI 110
Db 1736 NKEMKNTYNEFMESYNIIVLYLEFVSKESTYGEIKKRISTOKELLSIENVK----- 1790
QY 111 GKKIYIDGDTTKHLSLEALSEDKKIKIDYGRKALLHENVHVAKEGEVY----- 161
Db 1791 AKSYL--DDIETNEPRIRIYHFKKKINDVNDK--FTNEYSKVKNKGDNISINNVK 1844
QY 162 -----LVIOSSDEVENTER-----ALNVEYIGKILS----- 189
Db 1845 STDENLLSLINQTKENYANIVSKKYKYEAKENIFINISKLANSUNIQIONSSGIDLH 1904
QY 190 -----RQILS-----KINOPYOKFLVYLVNTIKN-----ASDSOG 219
Db 1905 KNINIAIVSYLNSQKEDMLSEIPSPQKTSSEYTKISDSYNTLIDIFPKSKXELQKKEOAL 1964
QY 220 DLFP-----TNOLKEHPDF-----SVFLEQNSNEVOEYFAKFAVYIE 259
Db 1965 NLITENRLHDQVATNELKQDLSDLKNNKQOILNKYXLLHKSNEILKSCNS----- 2018
QY 260 PQRDVLQVAPFAFNYMDKFNQEOINLSLEEL-----DQRMLSEYKMERIKQ 309
Db 2019 -QNYDTIL-----ESSKY-DKIKESNNVEQEKNNLGIDFVYAMEKKNNDIKIEELEN 2072
QY 310 HYQWSD-----SLSEBGRLL--KQLQ-----IPIEPKDD-----IHSLSQEK 349
Db 2073 NYKH-SDKDNYNFSEENNNILOSRRKLELSNAFNAEIKKEDKIERKNGLINLIETRK 2131
QY 350 ELL-----KRIQI-DSSDFLSTE--EKEFLKQLQIDISLSEEEKEL-----L 390
Db 2132 ECFMFTYTLVEALRIRITDYSKRTISATKFSKFLKIID-DTNSLNDIDNTLOTQYDL 2190
QY 391 NRIO-----VDSNPLSEKEKEFLKKLQIDQPYDINQRLQDTGGLIDSPSIND 440
Db 2191 NQIKKHTSMFADATINDNNNLIKEKEKATKTIINLTFL-----TIDSNIDAD 2239
QY 441 VRKQYKRDIONIDALLHOSIGS--TLNKKIYLYENMNINNLATLQADLYDSTON----- 493
Db 2240 VLHNNKIQMIYFENSELKSIDSIKQLYKKMHVFLNIGHINKKY-FDISKEFQNLQLO 2298
QY 494 -TKINRGIFNEFKK-NFKYSISSNYMIVDINERP-----ALDNERLKWRIOLSPDTR 543
Db 2299 ESELKENL-NDLKKIGOKISKKKKFLNELSEIPFNFNLKELYHIVAKESIDELKN 2357
QY 544 AGYLENGKLLIQNIGLEI-----KDVQIIKQSEK 573
Db 2358 ITNDENDNITLYMDIITLKKMKVESILNFVYTYENDSNVIKQIOTDNNNDVSKIKDNLK 2417
QY 574 EYIIDLAKVYK-----SKIDTKIOEAOINLNOEM--NKALGLPK- 611
Db 2418 KTISSFQKILNKMLKIKAQFYDNNNINNVISTISQDVIDYKHKHSLQDYENELIEIQKS 2477
QY 612 --YTKLITFNHNRVANSIVESAYLILNEMKNNISIOS-----DLJKVTNYLVDSNGR 661
Db 2478 LEYIKKSTYIDRSQITRYVAPHIDYEQQTKIQNDPNKDEIDLOELVINYKE----- 2533
QY 662 FVFTDITLPN-----IAQYTHODEIYQVHSGLYVPESRSIL 700
Db 2534 ---SELKLPITNNKDNVTPILSRIDKYINLIKSEYNNNDVSNVAVKK----- 2579
QY 701 LHGSKGVELLNDESGFIHERGHAVVDYAGYLLDKNQ---SDLYTNKSKKFIIDFK----- 752
```

```

Db 2580 -----LEEDANSIIFRD-----LDTSQNNLNLIDKLNKLTIDDKKKOE 2618
Oy 753 -EESNLTSTYGTNE 766
Db 2619 IENRYNLQITNREE 2633

RESULT 3
ID 096133 PRELIMINARY: PRT: 1979 AA.
AC 096133:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE HYPOTHETICAL 237.7 KDA PROTEIN.
GN PEB0145C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shalton S., Mason T., Yu K., Fujil C., Pederson J.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
DR EMBL: AE001375; AAC71819.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1979 AA; 237745 MW; 5C6CD6307AEFD37 CRC64;

```

```

Query Match 6.4%; Score 264.5; DB 5; Length 1979;
Best Local Similarity 19.0%; Pred. No. 0.0028;
Matches 169; Conservative 178; Mismatches 286; Indels 255; Gaps 40;

```

```

Oy 42 MHVKEKNKDNKRKDEERNKTOEHLKEIKHIVKIEVKEEAVKKBAEKLKLVPS 101
Db 329 LNKOKEREKEREREKEREREKEDTLIK-----ELKDEKI-----SILEKVS 376
Oy 102 DYLEHYKAIIGCIYIVDDITK--HISLEALSDDKKIKIDYKQALLHEHYVYAKEGE 159
Db 377 -----IKVREMDIEREHNFLH-MEDQKLDKNSFVKNN-NQLVYVCE--- 418
Oy 160 PVLVIOSSDEVVENTEKALNVEYEIGKILSRDILSKINQPYOKFLDVLNTIKNASDQ 219
Db 419 ----IKNKTELEKKEKELK---DIENVSKEIKLNKLNKEKQIILAFKNHKEE--- 468
Oy 220 DILFTNOLKEHPTDFSVFLEQNSNEVOEFKAPAYITEPOHRDVLQLYAEAFNYMDK 279
Db 469 ----IHGLKEELKE-SVKITKIEIETOELQEM-----VDIKQELDQL-----QEK 507
Oy 280 FNEOEINLSLELKQOHLMSRY-----EKMEIKQHYQWMSLSSE----- 321
Db 508 YNAQIESISIELSKEKEYNOKNTYIEIINLNKLEETNEKYLNQNNYNEINMLNN 567
Oy 322 ----GRGLKRLQIPIEPKDDI-----IHSLSOEKEL---LKRIOSSDPLSTEEK 368
Db 568 DIHMLNGIKTNTOISTLKNVHLNEQIDKLNNEKGLNKSISELNVQIMDL--KEEK 625
Oy 369 EFLKQLIDIRDSLEEEKELLNRIQVSSNPLSEKEKEFLKLKL----- 414
Db 626 DFLNQIYDLNSQI-----DLTRKMEKEKNMLQENKYNQEMELLRGNKISENILLN 680
Oy 415 DIQPDINDRL-----QOTGLIDSPSINLVDVRYOKYKRIQIINDALLHOS 459
Db 661 DEEVCDLKRKLSLSEKMKMKKEEHDKLAELKDCDVR--IRENNEKEDDINMLKEE- 737
Oy 460 IGSTLYNKIYLYENNNINNLATLGAADVSTDNTK-IRGIFNEPKKNFKXSIS-- 514

```

```

Db 738 -----YED-KIN-----TLKEQNEDKINTLKEQNEDKINTLKEEYEHKINTMKEE 781
Oy 515 -NMYIVDINERPALD---NERLKWRIQLSPTDRAGY-----ENCK 551
Db 782 YEHKINTLNQNEHKINTLNQNEHKINTMKEEYEDKMWNTLNQNEDKMNSLKEEYENKI 841
Oy 552 LILQRIQGLEINDY--QIIKQSEKEIYRIDAKVVPKSKIDTKIQEQLNLNQ----- 601
Db 842 NOINSNNEIKIDVNEVEYIEVDKLVTLDEK---KKQDFKEINVAHKAHEKEQIILTE 898
Oy 602 -EWNKALGLPKYTKLITFVHNRYASNIYESAVLIIN-EMKNQIOS-DLIKVTNLYVQ 658
Db 899 MEELKQQRNKKSDI-----YEKTI-KLKSICMIITIECCDIEDIEDIRIRIEYINN 952
Oy 659 NGRFVFTDITLPNIAQYHODEIYEQVHSKGLVPESSILHGPSKGEVLNDSGFT 718
Db 953 KGL-----KKEVEKEHKK-----HSSFNLKSEKF- 979
Oy 719 HERGHAVDDYAGYLLDKNOSDLVTNSKKF-----IDIFKEGSNL 758
Db 980 --FKNSIEDKSHELKKRHKEDLSKDEIEKNKKIKELINDIKLQDELIVYKKO-SNA 1036
Oy 759 TSYGRTEAEFEAEFRLMHSSTDHAERLKYOKNAPKTFQFINQIKFI 806
Db 1037 QQVDHKKKSWILLKDKSKKIDKENQINVEKNKEEDLKKKDEIRIL 1084

```

```

RESULT 4
ID 026216 PRELIMINARY: PRT: 2771 AA.
AC 026216:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE RHOPTRY PROTEIN.
OS Plasmodium yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5861;
RN [1]
RP SEQUENCE OF 379-2771 FROM N.A.
RC STRAIN-YM;
RX MEDLINE=97077455; PubMed=8920022;
RA Sinha K.A., Keen J.K., Ogun S.A., Holder A.A.;
RT "Comparison of two members of a multigene family coding for high-
RT molecular mass rhoptry proteins of Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 76:329-332(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-YM;
RA Green J.L., Holder A.A.;
RT "Structure of the E8 gene encoding a high molecular mass rhoptry
RT protein of Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 0:0-0(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-YM;
RA Holder A.A.;
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL: U36927; AAB41263.3; -.
SQ SEQUENCE 2771 AA; 325640 MW; C0CCB9A6E7ACF36 CRC64;

```

```

Query Match 6.4%; Score 263.5; DB 5; Length 2771;
Best Local Similarity 20.3%; Pred. No. 0.0047;
Matches 193; Conservative 168; Mismatches 358; Indels 233; Gaps 42;

```

```

Oy 2 NIKKEIKYISMCVLTATLTSPPVPIPLVQAGGAGDVGMHKEKEKKDKKRDDEER 61
Db 1759 NIYEPIKSYDILTHYLETVSKRPITYEQIKNKRITQAONELLTNINVKAKASYLDITFA 1818
Oy 62 NKTQE--EHUKETMKHI-----VKIEVKGE-----AKKKEAAKLEKVPDVL 104

```

```
Db 1819 NEPRRIYTHKRNKLVNDVNDKFTNEYSKVN-KGFPNISNINNVKKSDBENLLIINLQTK 1877
QY 105 EMKKAIGCKIYI-----VDG-DITKHS---LEALSEDK 134
Db 1878 EYMANIVSKKYSKYEAENIFINIPKLANSLNIQIKSSSGIDLFKNINIMILPYLDSOK 1937
QY 135 KKIDYIGKQALLHENVYAKEGEPEVL-VYQSESDYVENEKLANLYEIGKILSRDI 193
Db 1938 KDTLFTIPSEPTSETTYKISDSYNTLLDILKRSOELQKQOALNLIFF-----NRLH 1992
QY 194 SKINPOKQFLVNTIKNASDSODGLFTNOLKEHPDPSVEFLRONSNEVOEYFAKA 253
Db 1993 DKV-QAANELKQTLSDLKNKKEQ-----ILNKVK-----LLHKSSELKKSQNS 2036
QY 254 FAY--VTEPOHADVQLYAPAFNYMDKFNQEOINLSLELKQDRLMSRYEKMEKIKOHY 311
Db 2037 QNYDTILESSKYDKIK--EKSNNYEKEKEKLGINFQYKAMEQO-FNNIDIDIEKLENNY 2092
QY 312 QHMSD--SLSEGRGLL--KKIQ-----IPEPK---KDDIHSLSQEKKE-- 350
Db 2093 KHSEKDNVNFSEENNINLOSRRKLELTNAFNAEIKRIEDIIIEKNGLINLLETTRDCM 2152
QY 351 -----LLKRIQDSSDF-----LSTEEKEFLKQIDIRDSLESEKEL-----LNRI 393
Db 2153 LFTYKTLVEPLKIKITTYTKRITSATKFSKFEFLYIDA-TSNSLNDIOTLQTYKDLNQI 2211
QY 394 Q-----VSSNPLSEKEKEFLKKLKDIOPIYDINORLODTGGLDPSINLDVRK 443
Db 2212 NKHAVSVADATNDNNMLIEKEKEATKTINLTLEF-----TIDSNKIDADGLH 2260
QY 444 QYKRDIONDALHOSIGS--TLYNKIYVENMI-----NNLTATTLGADL 487
Db 2261 NKKIQTILFNSLHKSIDSIKQILKMKHAKRLNIGHINKRYEDISKEFDNILOLQSESL 2320
QY 488 VDSSTDKINKNGJFNNEKKNFKYSSISNYMIVDINERPALDNERKRIQLSPTTRAGYL 547
Db 2321 TANLNDKEIGQKISDKKQKQGLHALSEF-PIPNENTLKEIYHDIYVKKNOJDELENTNE 2379
QY 548 ENGLKILQIRNGLET-----KDVQIIOSEKEVYRIDAKVYPKSIDTKI 592
Db 2380 ENEMITLYMDITTKLMKVOSILNFVTYENDSNIIKOHIODNNENYVSKI-KESLETTYI 2438
QY 593 QEOALNIOENMKALGLPKYTKLITFPVNHRYASINIVESAVALLNEMKNOSLOLKKVT 652
Db 2439 QSFQKILN-----KLNKIKAOFTYDNNNNINNVISTISQDVNDVKKHISKDL--TIE 2486
QY 653 NYLVDSNGRFEVTDITLPIAEOY--THODEIYEOVHSGLYVESRSILHGPSKGYEL 710
Db 2487 NELIQ-----IQKSLIEDIKKSTYDIRSEQITKYVNPIDHYVQOQTKKIQNNPNK---- 2535
QY 711 RNDSEGIHREGHAVDYAGYILDKN-QSDL-----VTNSK-----KFIIDIRKEE 754
Db 2536 -----DEIDDLIOEIVYVKNESSELKPLTIINKNKNVTPPIISRDKVNLIKSE 2583
QY 755 GSN--LTSYGRNTEAEFFAEFRMLHSTDHAERL--KVQNAKPTQOFIND 801
Db 2584 YNNNDNYSYVAKLE--EDANNITIRDLTISHMLNDLIQKN-----KFIID 2629

RESULT 5
026223 PRELIMINARY; PRT; 2269 AA.
AC Q26223;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE RHOPTRY PROTEIN.
OS Plasmodium berghei yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5862;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN-YM;
```

```
RX MEDLINE-95021522; PubMed-7935623;
RA keen J., Slinh K., Brown K., Holder A.;
RT "A gene coding for a high-molecular mass rhoptry protein of Plasmodium
  yoelii."
RL Mol. Biochem. Parasitol. 65:171-177(1994).
DR EMBL: L27838; AAA21304.1;
SQ SEQUENCE 2269 AA; 265158 MW; F3D8CB103B9A6E1 CRC64;

Query Match 6.2%; Score 255; DB 5; Length 2269;
Best Local Similarity 20.5%; Pred. No. 0.0083;
Matches 210; Conservative 155; Mismatches 307; Indels 350; Gaps 51;

QY 39 DVGHNVEKEKKNKDNKRKDE-----ERNKQOEHLKEIMK---HIYKI---- 79
Db 1187 EVNSKIKNIETVQNHKKNYEIGIYERKINETAKTNKNOIESTKELIPTIOHIISSFNAN 1246
QY 80 EVKG-----EFAVK-EAEKLLKEVPSDVLNEMKKAIGKIYYDGD 120
Db 1247 DLEGIDSDENLGKYNTEGNIYEEFISYNLITLYLVSKESI--TYNQIQNKRIDTQKE 1305
QY 121 ITKHS-----LEALSED-----KKIKIDYIGKQALLHENVYAKEGEPEV-- 161
Db 1306 LKNIENNNKAKSYLDYIKENEPRIYTHKFKKINTV--NNEFKNE-YSKVNEGFDNISN 1362
QY 162 -----LYQSESDY--VENT-----EKALNYYEIGKI----- 187
Db 1363 SINTVKSTDENSLNLTNQTKEMVANNVTYYSKYEAENIFRNIPKLANSLNIKIN 1422
QY 188 -----LSRDI-----LSKINPOKQFLVNTIKNAS 214
Db 1423 SSGIDLSKDIKIALIYLDSTEDTLFIPSPQKTEYTIYSISYSLDILKSOELO 1482
QY 215 DSDGQ--DLF-----TNQLKEHPDF-----SVFLEQNSNEVOEYFAK 252
Db 1483 KKEQOTLKLIPENRRLVEKYQATNELGTLSDLYKKKEKILSEYKLLHKSSELNKLSCN 1542
QY 253 AFAYIIEPOHADVQLYAPAFNYMDKFNQEOINLSLELK---DQKLSYKEWME---K 306
Db 1543 F-----QNYDTIL--ESSKY-DQVKEKSNYKQEKELGIDFVNTQMEKEFNNDIK 1590
QY 307 IKOHQMSDSLSEGRGLLKKIQ-----IPIEPK---KDDIHSLSQEK-- 349
Db 1591 VIEELNENYDS-SEENNINLOSRRKLELTNAFNAEIKRIDDKIIEKNDLIDKLIETRN 1649
QY 350 -----ELKRIQDSSDLSTE--EKEFLKQIDIRDSLESEKEL-----LN 391
Db 1650 CMLFTHFTLAEFLKIKITDYKFIESAATKFSKEFLYIG-DTSNSLNDIATLQKYDLH 1708
QY 392 RIQ-----VSSNPLSEKEKEFLKKLKDIOPIYDINORLODTGGLDPSINLDV 441
Db 1709 QINKYVTSKLSDATNDNNMLIEKE-----ATQAIKNLTKLTIDSNNIDA 1755
QY 442 KROYKRDIONI--DALHOSIGS--TLYNKIYVENMI----- 476
Db 1756 NALNNKTIQMYFNSLHKSISTEIKQILKMKHAYKRLNIGINGKYDYSKQFONIIQLO 1815
QY 477 -NNLTATLGLADL--VSDTNTKINRGI-----FNEFKKNFKYSSISNYMIVDINE 523
Db 1816 ESELTANLN-DLKEIGQIKSDKKKKKFLHALNETPIPNENTLKEIYHDIYVKKRQIDEEN 1874
QY 524 RPAIDNE-----RLKRIQLSPTTRAGYLENGKLIQIRNGIIEKIKVOYIIOKSEK 573
Db 1875 ITSEENETITLYDTITRKLEKVOISILNFVTYENDSNIIKOHIODTENENVSXIKESLK 1934
QY 574 EYRIDAKVVPK-----SKIDPKIOEALNIOEW--NKAALGL 609
Db 1935 TTIOSFOEILKINGIKAQFIDNNNNINNNINNTISTISQDVNDVKKHISKDLTINELIEI 1994
QY 610 PKYTKLI--TFNV--HNRYASINIVESAVALLNEMKNNTIO--SDLIKVTYVLYDGN 659
Db 1995 QKSLIEDIKNSTYEIRGANNNNVNVTIRNVYEQQTKIQNNSKDEIDIDIQILNYYNKE-- 2052
```

QY 660 GRFVPTDITLPNIAEQYTHODEIYEOVHSGLYVPESRILHGPSKV-----ELRND 713
DB 2053 -----SETLPITGTGNKNNVTSTIISRI-NKVINLIESE-----YGNNNNVNNAKLEED 2102
QY 714 SGGFHEHGHAVDDYAGYLLDKNQS---DLVTNSKKFIDIFK-----EBSNLTYSYGR 764
DB 2103 ANSILID-----LDKSONILKDLIOQNLKTIIDDLKKNKQEIENRNMLQTINRE 2150
QY 765 NE 766
DB 2151 QE 2152

RESULT 6
Q9BX9 PRELIMINARY; PRT: 2747 AA.
AC Q9BX9;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE 235 KDA RHOPTRY PROTEIN (FRAGMENT).
GN ITA.2.
OS Plasmodium yoeII yoeII.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YM;
RA Khan S.M., Jarra W., Peter P.R.;
RT "Distribution and Characterization of the 235 kDa Rhoptry Multigene
RT Family within the Genomes of Virulent and Avirulent Lines of
RT Plasmodium yoeII.";
RL Mol. Biochem. Parasitol. 0:0-0(2001).
DR EMBL, AF323443; AAK15626.1; -.
FT NON_TER 1
FT NON_TER 2747
SQ SEQUENCE 2747 AA; 321237 MW; 430508AC7B8824BD CRC64;

Query Match 6.1%; Score 254.5; DB 5; Length 2747;
Best Local Similarity 18.8%; Pred. No. 0.011;
Matches 196; Conservative 168; Mismatches 298; Indels 379; Gaps 44;

QY 45 KEKENKDKENKRRKDEERNK-----TOEHLKE---IMKHIVKIVGGEAVKKEAAEK 94
DB 1648 KOKKNIEDOKKELDEVNSKIKNIENTVBQKKNYEIGIYEKINELAKTKNKNXI--ESTRE 1705
QY 95 LLEKVPDYLEMYKAIG-----GKIYIVDGDITKH-----ISLEAL 130
DB 1706 LKPTIOHTIISFNAXDLEGIDSDENLGKXTEGXNIYEETIKSYLITNYLETVESKE 1765
QY 131 SEDKKKIDYKDALLHMHYVAKEGYEPVLYIOSSEDEVTE-----KALNLY 181
DB 1766 TYNOJONKRIIDYOKELL-----KNIENVKAKSYLDYIKENEDRIVTHERKKKLNTV 1817
QY 182 YEIGKILSDISKINOPYOKFLDVNTIKNASDSGDGLFLTNCKE-----HPTDS 235
DB 1818 NDNFK-----NEKSKYNEGDNISNINIVKNSIDENSL-LNLTINCKEMANTIVNNTYS 1872
QY 236 VEFLEONSNEVOEFAK-----AFAYTIEPOHRDVL--- 266
DB 1873 YKY-----EAEINIFRNIPKLANTLNIKIKNSGIDLFKDIKALISYLSDKTEDTLFI 1926
QY 267 -----QIYAEAFNY-----MDKFNEOEINLSLE----- 290
DB 1927 PSPQKKTETTYTKISDSYSLIDILKKSQELQKKEOQTLKIFENRRLYERVOATNELRGT 1986
QY 291 -----ELKDORMLSRVE-----KWEKIKO---HYQWMSD 317
DB 1987 LSDLKTKKKEKIIISYVKLLHKSNLKLSNCFONTYTILESSKYDQVKEKSNNYXOEKEK 2046
QY 318 L-----SEGRGLKKLO-----IPIEP 335

DB 2047 LGIDFNVADEEKEFNNDIKYIEELNNVDSSEENNNILQSKOKLKLTLNFXAEIKIDD 2106
QY 336 K---KDDIHSLSQEEK-----ELKRIQDSDFLSTE---EKEFKLKIQIDI 378
DB 2107 KXIEKNDLIDKLITETKNCMLFTHTTIAETLKIKIDYSAFISARFSEFLKYIG-DT 2165
QY 379 RDSLSSEKEL-----LNRIQ-----VSSNPLSKEKEFLKLKLIDQPDIND 423
DB 2166 SNSLNDIATLQKLYDHQINKYVTSKLSDATNDNNLIEKE-----ATQ 2212
QY 424 RLDDTGLIDSPSINLVDKRYKRDIONI--DALHQSIGS--TLYNKIYENMNI--- 476
DB 2213 AIKNLTKLFTDSNNIDANLHNNKIQMYFENSELKSIESIQLKKNHVFLLNIGQI 2272
QY 477 -----NLTATLGADL--VDSDTNFKINGT-----FNEEK 505
DB 2273 NGKYPDISKOFDNILOLQSELTANLN-DLKEIGQKISQKNKFLHALNETPIPNFTLK 2331
QY 506 KNFKYSSISNVMVDINERPALDNE-----FLKMRIOISPRTAGYLENGKLIIQ 555
DB 2332 EYHDIVKTKRQIDELTENTISEENENITLYIDTITLKEKVSILNFTYENDSNITIK 2391
QY 556 RNIGLEIKDVQILQSEKEYIRIDAKVPR-----SKIDTKIOE 594
DB 2392 HIGDTNENDVSKIESLKTITQSFQELANKINGIKAQFYDNNNININNTISTQDND 2451
QY 595 AQLNIOEW---NKALGLPKYTKLI--TFNNINRASNIVESAYLILEKKNITOS--- 645
DB 2452 VKKHISKDLTXENELIEIQLSLIEDIKNSYVEISEOITVNTIRYVEOQTKIONNSN 2511
QY 646 -----DLIKVTYLVGNGRFVETDITLPNIAEQYTHODEIYEOVHSGLYVPESRIL 700
DB 2512 KDEIDIIQKILNKNK-----SETKLPITGTGNKNNVTSTIISRI-NKVINLIESE--- 2560
QY 701 LMGPSRGV-----ELRNDSEGFHEHGHAVDDYAGYLLDKNQS---DLVTNSKKFIDIF 751
DB 2561 -YGNNNNVSNVNAKLEEDANSIILD-----LDKSONILKDLIOQNLKTIIDDL 2607
QY 752 K-----EBSNLTYSYGRINE 766
DB 2608 KKKQEIENRNMLQTINREOE 2628

RESULT 7
Q9BK46 PRELIMINARY; PRT: 3130 AA.
AC Q9BK46;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE RETICULOCYTE BINDING PROTEIN 2 HOMOLOG A.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21101060; PubMed=11160005;
RA Trijila T., Thompson J., Caruana S.R., Delorenzi M., Speed T.,
RA Cowman A.F.;
RT "Identification of Proteins from Plasmodium falciparum That Are
RT Homologous to Reticulocyte Binding Proteins in Plasmodium vivax".
RL Infect. Immun. 69:1084-1092(2001).
DR EMBL, AF312916; AAK19244.1;
SQ SEQUENCE 3130 AA; 370415 MW; 13D973DB89D82026 CRC64;

Query Match 6.0%; Score 248.5; DB 5; Length 3130;
Best Local Similarity 19.8%; Pred. No. 0.023;
Matches 184; Conservative 146; Mismatches 274; Indels 325; Gaps 43;
QY 44 VKEKENKDKENKRRKDEERNKTOEHLKEIMKHIVKIVGGEAVKKEAAEKL----- 95
DB 1684 LKEKQNMVEAYKKMNMVNVNETE--KEIIKHKKNYEIRIMEHRIKETNEKKKKKFWESNN 1742

```

OY 96 -----LEKVPVLEMYKAIGKIIYVDGDTKHSLE 128
DB 1743 KSLTTLMDSPFRSMFYNEIYNDYNLNENPEKHQNLINELNGFNESYNILNPKMTEIINDN 1802
OY 129 ALSDDKKKINDI-----YK-----DALLHEHY---VYAKGEYEVLYIOSEDEVEMTEK 176
DB 1803 LDVYEIKEIKVQAEYEDKLNKKVDEL--KNYLNNIKEQEGHRLI-----DYK--EK 1851
OY 177 ALANYEIGKILSRDILS-----KINOPYOK-----FLDVLTNTIKNASDSGQDLF 223
DB 1852 IFNLXIKCSE--QONIIDDSYNYITVKQYIKTIEDVKFLDLSLNTIEKKNSVANLEIC 1909
OY 224 TNO-----LKEH-----PTDPSVEFLQNSNEVOEYAKAFAYIEPOH----- 262
DB 1910 TNKEDIKNLKHVIKLANFSGIIVMSDNTIEITPENPLEDDLL--MLQYFEKKHEITS 1967
OY 263 -----RDVQL--VAPFAFYMDKFN----- 282
DB 1968 TLENDSDLELDHLSNDESDIDNLKYNDITELHTTSTQLIKLYDNIQKLGDCNDLVKD 2027
OY 283 ---QEINLSLEELKDQRMLSRYEKWEKIKOHYOWSDSLSEEGRGLKQLQIPIEPBK- 337
DB 2028 CKELRELSTALYDLKIQ--ITSVINRENDISNNIDIVSNKLE-----IDAQYAFEKYKE 2081
OY 338 -----DIIHLSQEEKELKRIQIDS-----DPLSTEKEFLKQLQIDIRDSL 382
DB 2082 IFDNVEEYKTLDDTKMAYIKKABILKNVDINKREDLDIYFNDLDELEKSL-----TL 2135
OY 383 SEEEKELNRIQOVSSNPPLSEKEKEFLKLDIQPYDINORLODTGGLDPSINLQVR 442
DB 2136 SSNEMETKTIVQ--NSYNSFS-----DINKNIN-----DID 2164
OY 443 KOYKRDIONIDALLHQ--SIGSTLYNKIYELNNINNLATFLGADL--VDSTDN----- 493
DB 2165 KEMKTLIPMDELLENGHNDISLYNFI-----IRNIQIKIONDKINIREQENDRNIC 2217
OY 494 -----TKIRGIFNEKKKNFKYS--TSSNYMIVD-----INERPALDNER 531
DB 2218 FEYIÖNNYNFIKSDISIFNKYDHIKVDNYSNNIDVANKHNSLSEHVATNATNIEN-I 2276
OY 532 LKMRIOQSPDTRAGYLENGK---LILQNRIGLEIKDQVQIIKQSEKERYIDAKVYPPSKI 588
DB 2277 MTSIVEINEDTEMNSLEETQOKLELEYENF-----KKEKNITNNKYKIVHFKRL 2325
OY 589 DTKIQ---EAOININQENKALGLPKYTKLITFNVNHRKASINIVESAVALINENKNIOS 645
DB 2326 -KEIENSLLEYNSISTJFNFKI-----NETQNIDILKNEF--NNIKT 2363
OY 646 DLIKKYNYLVLDGNGRVPFDITL-----PNIAQYTHODEIYEQVHSK 689
DB 2364 KINDVK-----ELVHVDSITLLESIQTFNNLYGDLMSNIQDYKYEEDINNVELKKV 2415
OY 690 GLVYPPESRILLHPSPKGVELRNDSEGFIFHEFGHADVAYAG-----YL-LDKNOSDLV 741
DB 2416 KLYEINTNL-----GRINTFIKELDKYQDENNGIDKYIETINKENNSYI 2460
OY 742 TNSKKEIFDEKESGNTSYGRINAEFF 770
DB 2461 IKLEKANNLNKENSILLQNIKRNETELY 2489

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21101060; Pubmed-11160005;
RA Triglia T., Thompson J., Caruana S.R., DeIorenzi M., Speed T.,
RA Cowman A.F.;
RT "Identification of Proteins from Plasmodium falciparum That Are
RL Homologous to Reticulocyte Binding Proteins in Plasmodium vivax.",
DR EMBL; AF312917; AAK19245.1;
SQ SEQUENCE 3254 AA; 382876 MW; 6P9CAFA5AA6167BA CRC64;

Query Match 6.0%; Score 248.5; DB 5; Length 3254;
Best Local Similarity 19.8%; Pred. No. 0.024;
Matches 184; Conservative 146; Mismatches 274; Indels 325; Gaps 43;

OY 44 VKEKKNKDEKKRDEEKNKTOEHLKEMKHIYKIEVGEAVKKEAEKL----- 95
DB 1684 LKEQNNAEYKKNEMAYNYNETE--KEIHKKNYEIRIMEHIKKEETNEKKKFFMESNN 1742
OY 96 -----LEKVPVLEMYKAIGKIIYVDGDTKHSLE 128
DB 1743 KSLTTLMDSPFRSMFYNEIYNDYNLNENPEKHQNLINELNGFNESYNILNPKMTEIINDN 1802
OY 129 ALSDDKKKINDI-----YK-----DALLHEHY---VYAKGEYEVLYIOSEDEVEMTEK 176
DB 1803 LDVYEIKEIKVQAEYEDKLNKKVDEL--KNYLNNIKEQEGHRLI-----DYK--EK 1851
OY 177 ALANYEIGKILSRDILS-----KINOPYOK-----FLDVLTNTIKNASDSGQDLF 223
DB 1852 IFNLXIKCSE--QONIIDDSYNYITVKQYIKTIEDVKFLDLSLNTIEKKNSVANLEIC 1909
OY 224 TNO-----LKEH-----PTDPSVEFLQNSNEVOEYAKAFAYIEPOH----- 262
DB 1910 TNKEDIKNLKHVIKLANFSGIIVMSDNTIEITPENPLEDDLL--MLQYFEKKHEITS 1967
OY 263 -----RDVQL--VAPFAFYMDKFN----- 282
DB 1968 TLENDSDLELDHLSNDESDIDNLKYNDITELHTTSTQLIKLYDNIQKLGDCNDLVKD 2027
OY 283 ---QEINLSLEELKDQRMLSRYEKWEKIKOHYOWSDSLSEEGRGLKQLQIPIEPBK- 337
DB 2028 CKELRELSTALYDLKIQ--ITSVINRENDISNNIDIVSNKLE-----IDAQYAFEKYKE 2081
OY 338 -----DIIHLSQEEKELKRIQIDS-----DPLSTEKEFLKQLQIDIRDSL 382
DB 2082 IFDNVEEYKTLDDTKMAYIKKABILKNVDINKREDLDIYFNDLDELEKSL-----TL 2135
OY 383 SEEEKELNRIQOVSSNPPLSEKEKEFLKLDIQPYDINORLODTGGLDPSINLQVR 442
DB 2136 SSNEMETKTIVQ--NSYNSFS-----DINKNIN-----DID 2164
OY 443 KOYKRDIONIDALLHQ--SIGSTLYNKIYELNNINNLATFLGADL--VDSTDN----- 493
DB 2165 KEMKTLIPMDELLENGHNDISLYNFI-----IRNIQIKIONDKINIREQENDRNIC 2217
OY 494 -----TKIRGIFNEKKKNFKYS--TSSNYMIVD-----INERPALDNER 531
DB 2218 FEYIÖNNYNFIKSDISIFNKYDHIKVDNYSNNIDVANKHNSLSEHVATNATNIEN-I 2276
OY 532 LKMRIOQSPDTRAGYLENGK---LILQNRIGLEIKDQVQIIKQSEKERYIDAKVYPPSKI 588
DB 2277 MTSIVEINEDTEMNSLEETQOKLELEYENF-----KKEKNITNNKYKIVHFKRL 2325
OY 589 DTKIQ---EAOININQENKALGLPKYTKLITFNVNHRKASINIVESAVALINENKNIOS 645
DB 2326 -KEIENSLLEYNSISTJFNFKI-----NETQNIDILKNEF--NNIKT 2363
OY 646 DLIKKYNYLVLDGNGRVPFDITL-----PNIAQYTHODEIYEQVHSK 689
DB 2364 KINDVK-----ELVHVDSITLLESIQTFNNLYGDLMSNIQDYKYEEDINNVELKKV 2415
OY 690 GLVYPPESRILLHPSPKGVELRNDSEGFIFHEFGHADVAYAG-----YL-LDKNOSDLV 741

```

```
Db 2416 KLYINININL-----GRINTEFIKELDKODENNGIDKIEIENKENSNT 2460
Qy 742 TNSKFFIDFKEGSNLISYGRTEAEFF 770
Db 2461 IKLEKANNLKENFSKLONIKRNETELY 2489

RESULT 9
Q25662 PRELIMINARY; PRT: 1939 AA.
AC Q25662;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE REPEAT ORGANELLAR PROTEIN.
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5825;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=96V;
RA Werner E.B., Taylor W.R., Holder A.A.;
RL Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL; U43145; AAC63403.1;
SQ SEQUENCE 1939 AA; 229001 MW; B36E462001C6F22F CRC64;
```

Query Match 5.9%; Score 246.5; DB 5; Length 1939;
Best Local Similarity 19.6%; Pred. No. 0.016;
Matches 198; Conservative 174; Mismatches 302; Indels 335; Gaps 50;

```
Qy 36 GHGV--GMHYKKEKNNDENKRRKDEENKTOEHLKEIM-----KH---IKYIEYKG-- 83
Db 898 GHREYVAGLEEKHKKEVVALEEKHKKEELAKLEEGH--KEYMELGKHKKEVVAAGLAKHNL 956
Qy 84 EAVVKEAEKLEKVPDVL-----EMYKAIGKIIYVDGDTK-----HISLELSEDKK 135
Db 957 EEGHKEMVAE--LEKRHADLVAVLEEGHKA-----EIKIGEEKKEVVAAGIEEYK 1004
Qy 136 KIKDIYGDALLHEHYVAKEGEYVAVLIQSSSEYVENTEKALNYYEIGKILSRDILSK 195
Db 1005 KVEAL---KLAEKH---KD-----VYTKLEQHKKEELAKLEGGHKEVVAENEKKNASL 1051
Qy 196 INOPYOKFLDVNTIKNASDSGODLLFTNOLKEHPTDESVEFLDONSNEVOEYAKAFA 255
Db 1052 LAMLEENHKNEKIKLEEHKESASDLVERKLYOKDEEVKSNKKIEELTNVINDL----- 1105
Qy 256 YIEFQHRDVLQYAPFAFNVDKFPNE--QETN--LSLEELKD----- 294
Db 1106 -----NDSIMCYKKQILLEVEKKNENEELINKLIYONENKMDKKILEKENIKKL 1158
Qy 295 QRMISRYKMEKIKOHYH-----WSDSLSE---EGRGILKKIQPIE 334
Db 1159 NKKLSNYVFETKENTYKNSSEVYVNEKERILIVDSVCKENISESDVEGGKMLKMTLSLK 1218
Qy 335 PKKDDIHSLSOEKELLKRIQIDSSDLSTPEKEFLKLOI----- 376
Db 1219 -KKEKNITSINDKNK-----SSELVDITKSAYINKIEYKKEIEDNGKNIEDLKNNK 1269
Qy 377 -----DIRDSISEEKELLNRIQVDSNPILSEKKE---EFLK-----KTLK 414
Db 1270 ILDSLSELINLENKKNVLTDENNNKKLEIEI--KDKLNKEKENENMTIELINLNDIILKLLK 1328
Qy 415 DIQPYDINQRLQDTGGGLIDSPSINL-----DVRQKYKRDIONIDALLHQSIG--STLYNK 467
Db 1329 EISEW-----KDEEKLTKENIKLKNDIEQINKKEYKIKKEELMKIFMENINEVSLKNQ 1382
Qy 468 IYL-----YE-----NANI--NNLTATIGADLVDSYDN--TKINRIFNE- 503
Db 1383 IEIEEMKLEELKNYTELLAEKRETNMSISDNDKNIVENNILLDSKONNNKNKNVEEKT 1442
Qy 504 -----FKKNFYSSISNMIYDINERPALDNERLKMRIQLSPDTRAGYLNGKLLIQRNI 558
```

```
Db 1443 GDDINCEKNNDQAKETISYLKDELTKISMLXGEEL-----NRKNSYDEKAK--NL 1489
Qy 559 GLEIKDVQI-----IKOSEK-----EYIRID 579
Db 1490 TNEIKELKIRNKGGEAEIALLENLKNKIKERKNKSXKNDSSSSNNITTKDGKTRPEVSD 1549
Qy 580 AKVVP--KSKIDTKIOEADLINQENKALGLPKYIKLITFNVHNKNASNIYESAVLIIN 637
Db 1550 DKIQDKMKANLVKLKEKP-----DLMDNINSLEK-----ENRVSIVYKE----- 1590
Qy 638 EKMKNISDLIKKVTNYVLVDGNGR-----FVFTDI--TLIPNIAEOYTHODEIYEYOH 687
Db 1591 --NKNVQNKIYGIYISYFKKCEKELKNDMLVYICLVAKDLISILFLDNNVNLFEKIDKIL 1648
Qy 688 SKGLVVP--BSRSTLLHGSPGVELRN---DSEGFI---HEGNAV-----D 726
Db 1649 WKQWYIPTEIRILFLRYFSFLDKLRNYKCVNEEYVNNRERYESMALFQYLETASNKK 1708
Qy 727 DYAGYLLDKNOSDLVYNSK-----KFIDI-----FREEGSNLT 759
Db 1709 EMIIYVLEAEKDCSCENSSNEDKPKITDILNFSKDSIRLKTIAQLRKELNFEREAKNLL 1768
Qy 760 SYGRTEAEFFFAFRLMHSTDBAERLKYOKNAPKTFQF--INDQIKETII 807
Db 1769 NYDYQIILKKHYECLR-----KLKIVKNMARLEDPNIVNYSKFSI 1808
```

RESULT 10
Q9YV76 PRELIMINARY; PRT: 1127 AA.
ID Q9YV76;
AC Q9YV76;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ORF MSY156 HYPOTHETICAL PROTEIN.
GN MSY156.
OS Melanoplus sanguinipes entomopoxvirus (MSEPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OX NCBI_TaxID=83191;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TUCSON.
RX MEDLINE=99102612; PubMed=9847359;
RA Alfonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
RT "The genome of Melanoplus sanguinipes entomopoxvirus";
RL J. Virol. 73:533-552(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TUCSON.
RA Alfonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF063866; AAC97677.1; -
SQ SEQUENCE 1127 AA; 134265 MW; F185DDA1D5A3FE7D1 CRC64;

Query Match 5.9%; Score 245.5; DB 12; Length 1127;
Best Local Similarity 20.6%; Pred. No. 0.0088;
Matches 167; Conservative 126; Mismatches 263; Indels 253; Gaps 37;

```
Qy 60 ENRKTQOE---HLKELMKIIVIEYKGEAVVKEAEKLEKVPSPVLEMYKAIGKIT- 115
Db 24 ENNKVSLDINSLEYETLNI-----KFSDKITWEIINKVNYKIEYKIFV 66
Qy 116 -----IYDGDITTKHISLELSEDKKKIKD-----TYGDALLHEHYVAKEGEYVP 161
Db 67 MHNQKINDYNIILQYLIFENNEINKCTENKFPCKNPPL-----NITTKKKIYIID 118
Qy 162 LVIQSSSEYVENTEKALNYYE---ICKILSRDILSKINOPY-----QKFLDVL 207
Db 119 L-----DVEEKKDKELVINIQKNAVDKI--NDIKNNVNNHISDNETIITGKETLIDIL 170
```


DE PF3377.
 OS Plasmodium falciplarum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=96360472; PubMed=8719156;
 RX Handman E., Osborn A.H., Symons F., van Driel R., Cappai R.;
 RT "COS cell expression cloning of Pf3377, a Plasmodium falciplarum
 RT gametocyte antigen associated with osmophilic bodies."
 RL Mol. Biochem. Parasitol. 74:143-156(1995).
 DR EMBL; L04161; AAC37257.1;
 DR InterPro; IPR001064; Cystealin.
 DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
 SQ SEQUENCE 3119 AA; 377358 MW; 519B9D25BDEFCC CRC64;

Query Match 5.8%; Score 239.5; DB 5; Length 3119;
 Best Local Similarity 19.9%; Pred. No. 0.055;
 Matches 182; Conservative 153; Mismatches 313; Indels 265; Gaps 46;

OY 27 FTRPVGAGGHD-----VGMHVKEKKNKDNKDEERKNTQSEHLKEIKHIVKIE 80
 DB 964 FTTISGSEENHNRRLRKKEANLKEEMKRFNEQOEQERKKKAE--DENNETIQKH 1021
 OY 81 VKGEAEVKEAEKLEKVPDYLEMYKAIGKIVYDQ---DITKHISLEALSD--K 134
 DB 1022 METSKIEKEEVDYQDEFP-----KQYADQOELEFLTRDSEGESDVKR 1071
 OY 135 KKIK-----DIYKDALHEHYVYAEYEPVLVIOSSDYV-ENTERALNVEYET- 184
 DB 1072 DKVKPPRDPSPDFYNNALISFHEKMEELYN--TSISSLNYKEINRKFDDYVKEIK 1129
 OY 185 -----GKLSDILSKINQ-----PYQ-----KFLDVLNTIKNADSDG 218
 DB 1130 SKTYPRFEDLTQOTKTNCKKLFQKLETKDEYOKNIOGYKNKVIDLDDIQ--KKANG 1187
 OY 219 QDLFTNQLKEHPTDFSVFLEONSNEVOEFAKAF----- 254
 DB 1188 KTIILIONLIEKIDIKGVNRLSDRKFKKNFKVLRKKRMKMLDQFRAQFKAIRFINQ 1247
 OY 255 --AYIEPQHRDYLQ-LYRPAFNMYDKFNEQENISLEELKQRMLSR--YEKKERIK 308
 DB 1248 LTTTVEEETVLYEDIVYEKKRYKEEYS-----KMRRIISSMLDYEANKQIK 1296
 OY 309 OHYOHMSDSLSEEGRLKKIQIPIEPKKDDIHLSOEKELKRIQIDSSPFLSTEER 368
 DB 1297 EHY-HKVDITISEH-----KFOEIRQHMMDKIENITHELYKEMYVQIQIDLTNYV- 1344
 OY 369 EFLKLQIQRDLSSEKELNRIQVDSNP-----LSEKEEFLKRLQIDPYDIN-- 422
 DB 1345 -----HOLENHSELLQALQONKNIPHMLVLEKKELEITRRKKKN-KP-DISTS 1391
 OY 423 -----QRLQDTGGLIDSPSINLVRKQYKRQIDONIDALLHOSIGSTLYNKIYLYENM- 475
 DB 1392 SHATDEQVSDT--LIRGAHNGDIIKGEDND---EVLLIEQIQSL--KTMGQNOQ 1442
 OY 476 -----INNLATLGADLVDS-----TDWTKINRGJFNEKKKFKSISSNYIVINER 524
 DB 1443 VGSILEKLNLS-----DOYQLQDKRLNVEDIYKML-RNFKHYIEKLHESKINRE 1493
 OY 525 PALDNERLKMRIQLSPDTRAGYLENGKLLORNIIGLEIKDYQ-----IIKOS 571
 DB 1494 KFI-----TAVDULSNVYSTLEIYVVKFLHLDFQMSFEKDELEKHLLEYLE 1539
 OY 572 EKEYIRIDAKVVPKSKIDTKIQEAQ--LINOENMKALGPKYTKLITFNHNRVYASNI 628
 DB 1540 RKKYITLIEIQI--RDPLSTNIOQEGDHINNNNNN-----NVANNLKQY 1583
 OY 629 VESAVYLILMEKKNTI-QSOLIKYV-----NYLVDGNGRVPYDITLPNTAEQYTHODEIT 663
 DB 1584 LADDLLEISKLGHILEVCIKKNIALQOINYLNN-----TNETVPDVIKRLMAPRIV 1637

OY 684 ---EOVHSGKLVYVESKILLHGPSKGVLELNDSEGFHFEGHAYVDYAGYLDDKNOSDL 740
 DB 1638 PVSEDIYDITVWRDNTAV-----INNLTRHVRVMPFDQKIYDDHLL-----F 1681
 OY 741 VTNSKKFIDIFKE-----EGSNLTSGRTNE-----AEFFAEARLHNSDHAERLQKN 791
 DB 1682 VYNIKEL--TYKENLADKRYANYYEMNRFYLHLEEFF---YILKH---YVELRKIQDL 1733
 OY 792 AKPTQFINDQIK 804
 DB 1734 GEVAIPSEEBNIK 1746

RESULT 13

O9EMP3 PRELIMINARY; PRT; 1238 AA.

AC O9EMP3;
 DT 01-MAR-2001 (TEMBLrel, 16, Created)
 DT 01-MAR-2001 (TEMBLrel, 16, last sequence update)
 DT 01-MAR-2001 (TEMBLrel, 16, last annotation update)
 DE AMV156.

GN AMV156.
 OS Amsacta moorei entomopoxvirus (AmePV).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
 OC Entomopoxvirus B.
 OX NCBI_TaxID=28321;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20396580; PubMed=10936094;
 RA Bawden A.L., Glassberg K.J., Diggins J., Shaw R., Farmerie W.,
 RA Moyer R.W.;
 RT "Complete Genomic Sequence of the Amsacta moorei Entomopoxvirus:
 RT Analysis and Comparison with other Poxviruses."
 RL Virology 274:120-139(2000).
 RN [2]

RP SEQUENCE FROM N.A.
 RA Bawden A.L., Glassberg K.J., Diggins J., Shaw R., Farmerie W.,
 RA Moyer R.W.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF250284; AAG02862.1;
 SQ SEQUENCE 1238 AA; 147096 MW; 9BFB0A39DB6E35B CRC64;

Query Match 5.8%; Score 238.5; DB 12; Length 1238;
 Best Local Similarity 21.0%; Pred. No. 0.019;
 Matches 189; Conservative 152; Mismatches 276; Indels 281; Gaps 49;

OY 56 RKDEERNKQOEHLKEIKHIVKIEYGEBAVKEAEKLEKVPDYLEMYKAIGGI- 114
 DB 158 RSDQEFNINIKNYIHDIKN-----DNNIIDQLISRPIIDIIKKEFIDKLT 205
 OY 115 -----YIVGDITKHIS-----LEALSEDKKIKYIGKALLHEHYV---AKEG 157
 DB 206 NYIKNYISDKLVIINISKNPLFNINIKIYENINIKINNIINIK-IVNDDFLYFINNYKN 264
 OY 158 YEPVLVIOSSDEVEMTEKALNVEYEIGKILSD-----IISK-----INQPYOKF 203
 DB 265 YDFINLKNNEEDTIFLKNITINDY--INNLTSSDEYKNIITDMETIISKNIIDIMDYENK 323
 OY 204 LDVNLNTIKNNSDSG--QDLFTNQ--LKEHPDVSVEFLEONS-----NEVG 247
 DB 324 SDVIKKLKQKISDDEVYLSNLTSENENFLKKIKIISSENIDISLTNSLNINENKFKILK 363
 OY 248 EVFAKAFAYIEPQHRDYLQ-----YAPAFNYMDKN--EOETLSLEELK--- 293
 DB 384 EIINKEFL-----KGNIDMLVNNFNNDYKQKSIDEDIKIDINIKODEITHTLSISDLVI 438
 OY 294 -----DORMLSRYEKWEKIKQHYOHMSDSLSEEGGLLKL 329
 DB 439 SPYQNLFTYNDIDTMIESTYDKYDKQILNLVDETEKLPHHYKQVANSRPRQ---LSDV 494
 OY 330 QIPIEPRKDD-----IHSLSQEE-----KELKRIQIDSSPFLSTEERFLKIQ 375

D 495 SNSIDEKYNFIKNSALINNKINDKESFYFNVEDISKINILTNDDL-IFENNIENLN 553
Q 376 -----IDRD-SLSEEEKELNRIQVSSNPSEKEFEK-----FLKIDIQPD 420
D 554 SGCVTLNLDIKNIDSEIKIFINKI-----LPELIKFNLCGEVNSDKFNINNSV 606
Q 421 INQRLQDQGLIDPSILD-VRKQYKRDIONIALHQSIGSTLYN---KIYIEMMN 475
D 607 IN-----NMTDYILNLIKSNIPENSIL-TKVESRLKNYIDKFI-----D 647
Q 476 INNLATLADLVSDTMTKINRGIFENFKKFKYSISSNW---IVDINERPALDNRL 532
D 648 INNSISLNTNTOI--STNAKISE-IAADIEK-INKNISENKTIPIDEPDEKAKIDKIL 703
Q 533 KWRJQLSPDRAGYLKGLILQRLNIGLEIKQVOIIOSEKEYIR-----IDAKVVPK 585
D 704 K-----YLEIPVLQSKYA-----NIKEYIKSYTSRASLQIKNII 740
Q 586 SKIDTKIOEALNINOEN-----KALGLPKYTKLITFNVNNRNASNIVESAVIL 636
D 741 NILKRIISDNKINIEESNPSLSDKILKSIDLKLNQ-----DEIKE---L 786
Q 637 NEMKNINIOSDLIKKYTNLYLVGNGRFVFTDILP-----NIAEOYHODEI-----Y 683
D 787 SKTTEISSEDEKIN-----AINELLPLLESKNENIAGIY---DEINISMKY 833
Q 684 EQVHSKGLYVESRSILHGPSKVELRNDSEGFIERGHAVDYAGYLD-----NQ 737
D 834 NEILTK---IDNMVYLOE-----EKIENVLNINIDYFINVIDIKNKLKEINFIYKYN 885
Q 738 SDLVNNSKFFIDFKESGNTLSYGRTEAEFFAFLMSTDAERLAKYOKNAKPT 795
D 886 SQ--GNIPKYL---TEENTLRGT-----AETLNKIITYINSLENAPIINT-FSPVPT 932

RESULT 14
ID 026023 PRELIMINARY: PRT: 1048 AA.
AC 026023;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DS HYPOTHETICAL PROTEIN (FRAGMENT).
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=56329;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=3D7;
RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; M69147; AAA74653.1; -.
KW Hypothetical protein.
FT NON_TER 1 1048
FT NON_TER 1 1048
SQ SEQUENCE 1048 AA; 126518 MW; 7AF051480FA22424 CRC64;

Query Match
Best Local Similarity 19.8%; Pred. No. 0.022;
Matches 174; Conservative 149; Mismatches 301; Indels 254; Gaps 43;

Q 27 FIPVLQAGAGGCD-----VGMHVEKEKNKDNKRKDEERNKQOEHLKEIMKHIVKIE 80
D 295 FITISEGEENIDNRELKRKIKIANLKEWKKRFNEQOEQRERKKAEE--DEMNETIQKH 352
Q 81 VKGEAVKKEAEKLEKVPDVLWKYKIGKIYIVG-----DITKRIISLEALSSED--K 134
D 353 MESTKLEKKEVEDEVTODEED-----KQYAIQDQELEFLYLRTRDSEGSSESDVPK 402
Q 135 KLIK-----DIYGDALHEHYVVAKEGEYEPVLVIOSESDY-ENTERALNYYEYI- 184
D 403 DKVAFPDGSRPDSFYNTAISFHEKMEELYN--TSISSLNTYKELNKRFDVYKELK 460

Q 185 -----GKILSRDLISKINQ-----PYQ-----KFLDVLNITKNASDSOG 218
D 461 SKTYPKFDDLTSGFTKNCNKLFLQKLNFTIKDEYQOKNIQSYKKNKVIDILDIO--KKNAG 518
Q 219 QDLFTNOLKHPDVSVEFLQNSNEVOYFAFAF----- 254
D 519 KYIIONLILKIDIKYKGDVRLSDRKFKYKFRKYLGRKKMKLEDFRQKKAIRFIKD 578
Q 255 --AYIEPQHHRDLQ-LYAPAFNMDKFNQCEINLSLEELKDQMLSR---YEKWEKIK 308
D 579 LTTYMEETRYKVLIEDIYMEKKRYKEYS-----KMRRIISNLDIYENKOIK 627
Q 309 QHYQWSDSLSEEGRLIKLQIPIEPKDDI IHSLSQEEKELKRIQIDSSDFLSTEEK 368
D 628 EHY-HKVDITSEH-----KFOEIRQHRDKIENTIHELAKEMYVOIIDLTVNY----- 675
Q 369 EFLKRLQIDIDSLSEEEKELNRIQVSSNP-----LSEKEFEKLUKLQIDOPDIN-- 422
D 676 -----HOLENIHSELLQALQOKNIPRLNLVLEKRLKLETKRRKKN-KP-DISTS 722
Q 423 -----QRLQDTGGLIDSPSINLDVRKQYKRDIONIDALHQSIGSTLYNKIYIYENMN- 475
D 723 SHADQEQVSDT--LIRAHNHGDIKGEDND-----EVLLEIQLSL--KTKMDQNOQ 773
Q 476 -----INNLATLADLVDS---TDNTKINRGIFENFKKFKYSISSNVMIVDINER 524
D 774 VGSILEKLNLS-----DOYQLQDRLNVVEDIYKNTL-RNFKHYLEKHKESKINRE 824
Q 525 PALDNERLKWRIQLSPDRAGYLKGLILQRLNIGLEIKDQO-----IIKOS 571
D 825 KFI-----TKVDVLSNVYSTLEKWKFLHDOEWSFEKDELEKHYLEE 870
Q 572 EKEYIRIDAKVVPKSKIDTKIOEAQ---LNIQENKALGLPKYTKLITFNVNNRNASNI 628
D 871 RKKYITTEIQT--RDTLSTNQNQNGEGDHIINNINN-----NVRNNLKKQV 914
Q 629 VESAVLLINEMKNNI--QSDLIKKYT---NYLVGNGRFVFTDILPNIQAOYHODEIY 683
D 915 LKLDLLEISKLGHLEVDIKKIALEQINVLTNV-----TNETVPVVIDLMPAPRIY 968
Q 664 ---EQVHSKGLYVESRSILHGPSKVELRNDSEGFIERGHAVDYAGYLLDKNSDL 740
D 969 PVSEDIYVITWVRDNTAV-----INNTLRHFWMTQDKIYVDHLLI-----F 1012
Q 741 VTNSKFFIDFKF-----EGSNLTSYGRTEAEFFAFAF 774
D 1013 VYNIKEL--IKKNLADKYNANYEEMNRFYLNHEEF 1048

RESULT 15
ID 077320 PRELIMINARY: PRT: 3724 AA.
AC 077320;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE PFC0335C PROTEIN.
GN PFC0335C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5633;
RN [1]
RC SEQUENCE FROM N.A.
RP Mungall K., Lawson D., Barrett B.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z98547; CAB1104.1; -.
SQ SEQUENCE 3724 AA; 448204 MW; 1D849821C4D8E904 CRC64;

Query Match
Best Local Similarity 19.1%; Pred. No. 0.19; Mismatches 352; Indels 218; Gaps 42;

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:54:01 ; Search time 132.69 seconds
(without alignments)
857.637 Million cell updates/sec

Title: US-09-747-521-2_COPY_1_778

Perfect score: 3987
Sequence: 1 MNKKKFFIKVISMCLVTAI.....TSYGRTEAEFAFAFRLLMH 778

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	273.5	6.9	2752	5	Q9BJY0 plasmodium
2	268	6.7	1946	5	Q97291 plasmodium
3	255	6.4	2269	5	Q26223 plasmodium
4	254.5	6.4	2747	5	Q9BJX9 plasmodium
5	253	6.3	1979	5	Q96133 plasmodium
6	253	6.3	2771	5	Q26216 plasmodium
7	248.5	6.2	3130	5	Q9BK46 plasmodium
8	248.5	6.2	3254	5	Q9BK45 plasmodium
9	245.5	6.2	1387	5	Q9G276 plasmodium
10	239	6.0	1127	12	Q9YV76 plasmodium
11	237	5.9	1238	12	Q9EMP3 plasmodium
12	235	5.9	1048	5	Q26023 plasmodium
13	235	5.9	3119	5	Q25857 plasmodium
14	233.5	5.9	1939	5	Q25662 plasmodium
15	228.5	5.7	1786	5	Q9U0P0 plasmodium
16	225.5	5.7	1931	5	Q9NCF9 plasmodium
17	224.5	5.6	806	2	Q9L8P7 mycoplasma
18	224.5	5.6	1558	5	Q96275 plasmodium
19	221.5	5.6	3724	5	Q77320 plasmodium

20	220	5.5	2166	2	Q51465 borrelia bu
21	218.5	5.5	1931	5	Q9VKH9 drosophila
22	218	5.5	980	5	Q96246 plasmodium
23	217	5.4	1088	5	Q00905 oxytricha f
24	216.5	5.4	2867	5	Q9N2M5 plasmodium
25	214.5	5.4	1313	10	Q9XIP6 arabisdopsis
26	213	5.3	1516	5	Q96154 plasmodium
27	211.5	5.3	1206	4	Q9NWC1 homo sapien
28	209.5	5.3	976	2	Q54222 staphylococ
29	209.5	5.3	1005	1	Q58718 methanococ
30	209.5	5.3	1128	10	Q9SAP6 gsaite arabidopsis
31	209.5	5.3	1780	4	Q9UFR5 homo sapien
32	208.5	5.2	999	5	Q9U5A3 plasmodium
33	207.5	5.2	683	2	Q50281 mycoplasma
34	207	5.2	849	2	Q9AHK4 borrelia bu
35	206.5	5.2	1650	5	Q77328 plasmodium
36	206.5	5.2	1734	5	Q9U0L7 plasmodium
37	205	5.1	1065	2	Q9AHK8 borrelia bu
38	204	5.1	1065	2	Q9AHK7 borrelia bu
39	203.5	5.1	2139	5	Q07569 entamoeba h
40	203	5.1	849	2	Q9AHL2 borrelia bu
41	203	5.1	3259	4	Q14789 homo sapien
42	202.5	5.1	1070	10	Q9AW67 guillardia
43	201.5	5.1	841	5	Q23037 arabidopsis
44	201.5	5.1	1909	5	Q25893 plasmodium
45	201.5	5.1	1938	13	Q91BD7 seriola dum

ALIGNMENTS

```
RESULT 1
Q9BJY0 PRELIMINARY; PRT: 2752 AA.
AC Q9BJY0;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE 235 KDA RHOPTRY PROTEIN (FRAGMENT).
GN IIA.1.
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YM;
RA Khan S.M., Jarra W., Peter P.R.;
RT "Distribution and Characterization of the 235 Kda Rhoptry Multigene
   Family within the Genomes of Virulent and Avirulent lines of
   Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 0:0-0(2001).
DR EMBL, AF323442; AAK15625.1; -.
FT NON_TER 1
FT NON_TER 2752
SQ SEQUENCE 2752 AA; 322606 MW; E834E3753168AF76 CRC64;
```

Query Match 6.9%; Score 273.5; DB 5; Length 2752;
Best Local Similarity 19.6%; Pred. No. 0.0017;
Matches 191; Conservative 158; Mismatches 283; Indels 343; Gaps 44;

```
QY 58 DERKNKTGEHKEIMKHVKIEVGEAV-----KKKAELKLEKVPVLEVKAI 110
DB 1736 NKEMNNTYFEMESYNLIVDYLETYSKESITYGEIKNKRISTOKELKIEENVK----- 1790
QY 111 GSKIIYVDGDIRKHISLEALSEDKKIKIDYGRDALHHEVYAKGEVPEV----- 161
DB 1791 -AKSTL--DDITNEEDRAIVTHFKKKINDVNDK---FTMEYKVKNGCFNINSINNVK 1844
QY 162 -----LVYQSEDIYVENTK-----ALNYYEIGKILS----- 189
DB 1845 SYDENLLSLILQTKEMVANIYSKKYYSKYEAENIFINISKLANSLNIOIONSSGIDLH 1904
```

```
QY 190 -----RDIIIS-----KINOPYOKFLDVLNTIKN-----ASDSGQ 219
D 1905 KNINIAIVSYLNSOKEDMSIFSPKQSTETKYKISDSYFTLLDIFKSKYELQKKEQAL 1964
OY 220 DLFF-----TNOLKEHPTDF-----SVEFEONSNEVOEFARAFAYIE 259
D 1965 NLIFENRILHDKVOATNELKDTISDLKKNKEQLNKVXLLHKSNELNKISCS----- 2018
OY 260 POHRDVLQVAPAFNPMKFNDOEINLSLEELK-----DORMLSRYEKWEKIKQ 309
D 2019 -QNYDTLL-----ESSKY-DKIKESNNYEQEKNKLGIDFVTAAMEKFNNDIDIELEN 2072
OY 310 HYOHMSD-----SISEGRGLL-----KKLQ-----IPIPKDD-----IHSLSOEK 349
D 2073 NYKH-SDKDNYNFESENNNLQSKKIKELSNAPNAEIKKXEDKIEKNKLIETTRK 2131
OY 350 ELI-----KRIQI-DSSDFLSTE-----EKPEFLKILIDIRDSLEEKEL-----L 390
D 2132 ECHFEFTYTLVLEALRIKIDYSKFTSATKFSKEFLKYID-DTNSLNDINDINTLQTKYDL 2190
OY 391 NRIO-----VDSNPLSEKEKEFLKLLKLDIOPYDINORLODPTGGLDPSIIND 440
D 2191 NOIKKHVTSMEFADATNDNNNLIEKEKATKTINNLELF-----TIDSNINDAD 2239
OY 441 YKQOYKRDIONIDALLHOSIGS--TYNKTYLYENMNINNLATLGCADLVDSIDN----- 493
D 2240 VLNHNKTOMIYFSELSKIDSIKOLYKKNHFKLLIGHINKRY-FDISKEPNDILOQ 2298
OY 494 -TKINGIFNEFKK-NFKYSSSNMYIVDNERP-----ALDNERLKWRIQLSPDTR 543
D 2299 ESLEKLENT-NDLKIKGOKISDKKNKFLNELSELPINPENTLKEIYHEIYVESQIDEIKN 2357
OY 544 AGYLENGKLIQNRIGLET-----KVOIIRKQSEK 573
D 2358 ITNDENDNITLYMDITIKLKKVYESILNFYTYENDSNVYIKOHIONNENDVSKIKINDLK 2417
OY 574 EYIRIDAKVPR-----SKIDTKIQEOQLMINQW--NKAIGLPR- 611
D 2418 KTESQOKIINKNETKAQFYDNNNNINNVISTISODVIDVKHHSKDLTYENELIEIORS 2477
OY 612 --YTKLITFNHNRVASNIVESAYLLIEMKNKNIQS-----DLIKKTYNVLVDGNGR 661
D 2478 LEVTKKSTDIRSEQITKYVNPIDHYVEQOTKKIIONPNKDEIDDLQELVYNKKE----- 2533
OY 662 FVETDITLPR-----IABQYTHODEIYEOVHSGLYVPESRSIL 700
D 2534 --SEKLPITINNKDNVYPIISRIDKVINLIKSEYNNNDNVSNVAKK----- 2579
OY 701 LHGRPSGVELRNDSEGFIEFGHAVDYGILLDKNQ---SDLVTSKKRFIDIFK----- 752
D 2580 -----LEEDANSTIRD-----LDTSONMLNDLILQNKLIIDDLKKNKQE 2618
OY 753 -SEGSNLTSGRTNE 766
D 2619 IENRXNLQITINRQE 2633

RESULT 2
ID 097291 PRELIMINARY: PRT: 1946 AA.
AC 097291:
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE HYPOHETICAL 231.8 KDA PROTEIN.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota: Alveolata: Apicomplexa: Haemosporida: Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=99376085; PubMed=10448855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
```

```
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jaggels K., Jassal B., Kyes S., McLean J., Moule S.,
RA Mangall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.,
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum .";
RL Nature 400:532-538 (1999).
DR EMBL: AL034559; CAB39037.2;
DR InterPro: IPR002048; EF-hand.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_2.
KW Hypothetical protein.
SQ
SEQUENCE 1946 AA; 231792 MW; 59AC248AB9808E34 CRC64;

Query Match 6.7%; Score 268; DB 5; Length 1946;
Best Local Similarity 21.5%; Pred. No. 0.0019; Indels 208; Gaps 43;
Matches 183; Conservative 153; Mismatches 306;

OY 47 KKKNKDEKKRDEERNKQ-----EHLKEIMKHIVKIEVGEEAVKR 89
D 1051 RDKNHNNSNNKNNKNNNNNNYVYQHNNNSHITVLEKKKALKOKYVLENNILVQKK 1110
OY 90 EAA-----EKLEKVPDVLNMTKAIQKTY---IYDGIYHISLEALSEDKKIKDI 140
D 1111 EMFECNKKERKKYKKKISLNEYEKKLDNIITDFNKLKENCITKEKLLAKMEDITKYINEQ 1170
OY 141 YGRDAL-----LNEHYVAKEGEYPLVIOSSPDYVENEKALNVYELKILSRILSK 195
D 1171 FSLSKIOFENKMDYVFLKKKDSIYWL--KELIEKKKTYLNDNLIKQYKKDDDI 1227
OY 196 INOPYOKFLDVLNTIKN---ASDSGQDILFTNOLKEHPTDSVEFEONSNEVOEFA 251
D 1228 LKENIEKIDIRKKLQKEEITISQKRIQETLENLKK-----IGKRIKFNDEIQKLOY 1282
OY 252 KAFAYIEPO-----HRDVLQVAPAFNPMKFNDEIWL--SLE-LNDORML- 299
D 1283 K-INIHEKETIEKTDNIEKERNKTKSNKFDIINKNNNLHNKIEITLNEKEISTENV 1341
OY 300 -----RYEKWEKIKOHYOHMSDSLSEGRGLKLIPIEPKK 337
D 1342 KLETKNTLYIENKLSNDLKNLKEKEKIQEN---KNINKNEHIIELOSYPEIKE 1397
OY 338 DDI-----IHSLSOEKELKRIQIDSDFLSTEKEFLKLIQIDIRDS--LSEEE 366
D 1398 KHKKEITQIEQOIHNLDKH---IAQIOIEKNF-----ESSYLKEKNENKMSNIIEKY 1449
OY 387 KEL-----LARIQV-----DSSNPLSEKEKEFLKLLKLDIOPYDINORLODPTGLID 433
D 1450 KELSTYEIDKNINKIKIEDLEKEKENILTKNE-INNLKEEYKM--VOQHLEDFTVLYE 1506
OY 434 SPISIND-VKQYKRDIONIDALLHOSIGSTLYNKIYLYENMN-----INNLATLGADL 487
D 1507 KOKLAIDITITKEKNNTINEDCKIKKNK--KLLNKK--LKENONNYEITLNNI----- 1554
OY 488 VDSNTNKTNRGIFNEKKNFKYSISS-----NY-MIVDINERPALDNERLKWRIQLS 539
D 1555 -----KKENOQIIEREKKNFTQKVESLEHAFKOSYQNLKDONENLQOQIKOLK-NVNO 1607
OY 540 PDRAGYLENGKLIQNRIGLETQVQIIOSEKEVYIRIDAKVPRKIDTKIQEOQLNI 599
D 1608 IKTNKSLKLVNELL-----IKETKNYSQOKEKFIK-GLKNIKQAYIKLKNENQDLKI 1659
OY 600 NO-EMNKALCLPKYTKLITFNHNRVASNIVESAYLLIEMKN-NIOSDLIKKVTNVLVD 657
D 1660 NAFEYIKKDVQDN---VTLNVHNN-----ILNOKKLFQVIDILKS---QVD 1701
OY 658 GNGREYFTDITLPAEQYTHODEIYEOVHSGLYVPESRSILLHGSKVVELRN----- 712
D 1702 OKONII-----NNMKEO-----IEDVNHKIASINKKEEL---NTTIKIKNKTTED 1744
OY 713 ---DSEGFIEFGHAVDYGAILDKNQSDLVTSKKRFIDIFNEGSNL-TSYGRTNEAE 768
```

Db 1745 VNLSEVKLSKSELSKDEVKRKRTIEKQEE--REYKKLLDDYKIEKKNLVTEYKE--EID 1800
QY 769 PEAEAFRLMH 778
Db 1801 STMYKEFAH 1810

RESULT 3
026223
ID 026223 PRELIMINARY: PRT: 2269 AA.
AC 026223:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-NOV-1996 (TREMBLrel. 08, Last annotation update)
DE RHOPTRY PROTEIN
OS Plasmodium berghei yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5862;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YM:
RX MEDLINE=95021522; Pubmed-7935623;
RA Keen J., Sima K., Brown K., Holder A.;
RT "A gene coding for a high-molecular mass rhoptry protein of Plasmodium
RT yoelii.";
RL Mol. Biochem. Parasitol. 65:171-177(1994).
DR EMBL: L27838; AAA21304.1; -
SO SEQUENCE 2269 AA; 265158 MW; F3D8CB103FB9A6E1 CRC64;

Query Match 6.4%; Score 255; DB 5; Length 2269;
Best Local Similarity 20.5%; Pred. No. 0.0081;
Matches 210; Conservative 155; Mismatches 307; Indels 350; Gaps 51;

QY 39 DVGHHVKEKKKKDKKKRDE-----ERNKTOEHLKEIKM---HIYKI---- 79
Db 1187 EVNSKIKNIENVTNQHKMYEIGIVEIKINEIAKTNKNOQESRKELIKPTIOHIISSFNAN 1246
QY 80 EVKG-----EEAVKK-EAAEKLEKPSVLEMYKAIIGKTIYVDCD 120
Db 1247 DLEGIDSENLGKYTEMGNIIYEEFIKSYNLITNTLETYSKESI--TYNOIQKRRIDTOKE 1305
QY 121 ITKHIS-----LEALSED-----KKIKDIYKGDLLHHEHYAAKEGYEPV-- 161
Db 1306 LKKNIEENNKAASYDIYIKENEFDRIVTHFKKKLNTV--NDNFKNP-YKVNDEGEPNIN 1362
QY 162 -----LVIOSEEDY---VENT-----EKALNVEYEIGKI----- 187
Db 1363 SINTVKNSTDENSLNLTINQTEKEMAYINVTNYYSYKAEANIIFRIPKLANTLNLIKIKN 1422
QY 168 -----LSRDI-----LSKINOPYQKRLDVLNTIKMAS 214
Db 1423 SSGIDLSKDIKAILSYDSKTEDTLIFPSQKTEYTYKTSIDSYSLDILIKKSQELQ 1482
QY 215 DSDGQ--DLLF-----TNOLKEHPDVF-----SVEFLQNSNEVOEFAK 252
Db 1483 KKEGOTLKLIFENRRRLYEYQATNELLRGTLSDLKAKKKEKILSEVALLHKSSEINKLSGN 1542
QY 253 AFAYIEPOHRDVLQVLAPEAFVMDKNEQINLSLEELK--DQRMLSRYEKME--K 306
Db 1543 F-----QNYDTIL-----ESSKY-DQVKEKSNNYKQEKELGIDFNVTVDMEKEFNNDIK 1590
QY 307 IKOHQHMDSLSSEGRGLKKLQ-----IPIEK-----KDIHISLSQEEK- 349
Db 1591 VIEELENNYDS--SEENNNTLOSQKQKLKELTJNKFNMAEIKKIDKIIKKNLDLKLLETNRN 1649
QY 350 -----ELKRIQIDSDFLSTE--EKEFLKKLQIDIRDSSEEEKEL-----LN 391
Db 1650 CMLFHTHTLAEFLKIKIDYKFIKSAFKFSKFLKAYIG-DISNSINDIATLQKLYDHA 1708
QY 392 RIQ-----VDSNPLSEKEKEFLKKLKLQIDQPYDINQLODTGGLDPSINLDV 441

Db 1709 QINKVYTSKLSDATNDNNNLIENEKE-----ATQAIKNLTKLFTIDSNUNIDA 1755
QY 442 RKQYKRDIONI--DALLHOSIGS--TLYNKIYXEMNT----- 476
Db 1756 NALHNKKIQMVFNSELHSIESIKOLYKRMHVKILLNIGQINGKYFDISKQFDNLQLO 1815
QY 477 -NNLFTATGADL--VDSNTNTRKINRGI-----PNEFKKNFYSISSNMYIVDINE 523
Db 1816 ESELTANLN-DLKEIGOKISDKKKNKFLHALNENPIFNFTKIEYIHDYKRRQIDEIEN 1874
QY 524 RPAIDNE-----RLKWRIQLSPTPRAGYLENGKLILQNLNLEIKDVOIIOKSEK 573
Db 1875 ITSEENENITLYIDITTKLEKQVOSILNFTYENDSNIQKHQIDPTNENDVSKIESKL 1934
QY 574 EYIRIDAKVPR-----SKIDTIQEQQLNINQEW--NKALQL 609
Db 1935 TTIOSEFELTKNGIKAKOPYDNNNNINNTISTISQVDNKKHISDQLTENLEIET 1994
QY 610 PRYTKLI--TFNV--HNRYASNIVESAYLLILEMKNINIO--SDLIKVTNYLVDSN 659
Db 1995 QKSLEDIKNSTYEIRGANNNYVTINRYEQQTINKQNNNSKDEIDDILOKILNTYKE-- 2052
QY 660 GRVFETDITLPNIAEQYTHODELYEYHSGLYVPSRSITLHGSKV-----ELRND 713
Db 2053 -----SETKLPTTGKNNVTSIISRI-NKVILIESE-----YGNNNNVSYNAKRLIED 2102
QY 714 SEGFHEPFAVDVAGYLLDKNOS--DLVTNSKKEFIDIFK-----EGSMLTSYGR 764
Db 2103 ANSTIID-----LDKSQNLTKDLQOOLKTIIDDLKKNKOEIERNNLOTINRE 2150
QY 765 NE 766
Db 2151 QE 2152

RESULT 4
09BJX9
ID 09BJX9 PRELIMINARY: PRT: 2747 AA.
AC 09BJX9:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 235 KDA RHOPTRY PROTEIN (FRAGMENT).
GN IIA.2.
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YM:
RA Khan S.M., Jarrar W., Peter P.R.;
RT "Distribution and Characterization of the 235 kDa Rhoptry Multigene
RT Family within the Genomes of Virulent and Avirulent Lines of
RT Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 0:0-0(2001).
DR EMBL: AF323443; AAK15626.1; -
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 2747 AA; 321237 MW; 430508AC7B8824BD CRC64;

Query Match 6.4%; Score 254.5; DB 5; Length 2747;
Best Local Similarity 18.8%; Pred. No. 0.011;
Matches 196; Conservative 168; Mismatches 298; Indels 379; Gaps 44;

QY 45 KEKKNKDKENKDKERKNK-----TOEHLKE---IMKHVYIEVKGEEAVKKEAAR 94
Db 1648 KQKNIEDQKLEDEVNSIKNIENVTNQHKMYEIGIVEIKINEIAKTNKXN--ESTRE 1705
QY 95 LLEKPSDVLEMYKAI-----GKIYVDGDIYH-----ISLEAL 130
Db 1706 LKPTIOHIISSFNAXDLEGIDSDENLGRKXXTEGNGIYEEFIKSYNLITNTLETYSKESI 1765

```
QY 131 SEDKKIKIDYKDALHEHYVAKGEPEVLYQSSSEYVENTE-----KALNVY 181
Db 1766 TYNDIQMKRIDTQKELL-----KNIEVNNKASLYDIYKENEFRIVYTHKKKINTVY 1817
QY 182 YEIGKILSRDILSKINOPYQKFLDVLNTIRKASDSDODLLFTQNLKE-----HPTDFS 235
Db 1818 NDNRK-----NEXSKVNVGFGFNDISINSITVKNSTDENSL-LNLTQTEGMANVANNYYYS 1872
QY 236 VEFLEQNSNEVQEVFAK-----AFAYIEPQHRDVL--- 266
Db 1873 YKY-----EAENIFRNIPKLANTLNIRKSSGIDLFKDIKIAILSYLDSKEDTLIFI 1926
QY 267 -----QVAPFAFNY-----MDKFNQEIINLSLE----- 290
Db 1927 PSPQKTEYTKISDSYSILLDIKKSQEQLQKQEQTLKLFERNRLRYEKVQATNELRG 1986
QY 291 -----ELKDQMLSRHY-----KNEKIKQ--HYQHSMS 317
Db 1987 LSDLKYKKKILSEVKLLHKSNLKNLSCNFQNYDTILLESSKYDYKESNNYXQEK 2046
QY 318 L-----SEEGRLIKKLQ-----IPEP 335
Db 2047 LGIDFNADMEKENNDIKVLEENNVDSSENNNIIQSKQKLELTNKFAXEIKRID 2106
QY 336 K-----KDDIHSLSQEEK-----ELIKRIQISSDFLSTF---EKPELKKLQIDI 378
Db 2107 KXLEKNQDLQKLETRKNQCMLEHTTTLAETLKIKITYSKTESATYFSEFLAYIG-DT 2165
QY 379 RDSLSSEKEL-----LNRIQ-----VDSNPLSEKEKEFLKLLKIDQIDYQ 423
Db 2166 SNSLNDIATLQKLYDHLQNKVYTSKLSDATNDNNLIEKEK-----ATQ 2212
QY 424 RLQDTGLDPSJNDVRYQKRDIONT--DALLHOSIGS--TLYNKIYLYENMNT--- 476
Db 2213 AIKMLTKLFTIDSNNIDANALHNKNIQMYVYENSELHKSIESIKOLYKMHVFKLLNTGQI 2272
QY 477 -----NNLTATLGADL--VDSDTNFKINRG-----FNBEK 505
Db 2273 NGKFEDISKOPDNILQOSELENTANLN-DLKEIGQISKDKNKKFLHALNETPIPNFTLK 2331
QY 506 KNFYSISSNMYIVDINERPALDNE-----RLKWRIQLSPTPRAGYLENGKILQ 555
Db 2332 EITHDIYKRYQKIDEININISEMENTIYIDITTKLKEVQOSILNVYTYENDSNIIKQ 2391
QY 556 RNIGLEIKDVOIKQSEKEYIRIDAKVPR-----SKIDTKIQE 594
Db 2392 HIQDINENDVSKIESKLTQIQSEFQILNKINGIKAQFYDNNNNINININTISTISQVND 2451
QY 595 AQLNINQEW---NKALGLPYTKL---TFNVNHRVASNIYESAVYLLNEKKNQIS--- 645
Db 2452 VKKHISQDLYXENBELIQSLQSEDIKNSYETIRSEQITNTVNTIRNVEEQDTNKIQQNSN 2511
QY 646 -----DLIKRVTVLVGNGRFVFTDILPNIAQYTHODEIYEOVHSGLYVESRSIL 700
Db 2512 KDEIDITQKILNVNKE-----SETKRLPTTGKNKNVYISIRI-NKYVINLSE--- 2560
QY 701 LHGPKSVY-----ELNDSBGFHIEGHAVDYAGYLLDKNQS--DLVYNSKKFLDIF 751
Db 2561 -YGNNNNVSYNAKKLEDEDANSILD-----LKSQWILDLIOQNKIIDDL 2607
QY 752 K-----EEGSNLTSYGRTNE 766
Db 2608 KNKKQELERNRLQITNREOE 2628

RESULT 5
096133 PRELIMINARY; PRT; 1979 AA.
AC 096133;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE HYPOHETICAL 237.7 KDA PROTEIN.
```

```
GN PRB0145C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tetteh H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallom S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pateria M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
DR EMBL; AE001375; AAC71819.1;
KM Hypothetical protein.
SQ SQUENCE 1979 AA; 237745 MW; 56C6D6307AF7D37 CRC64;

Query Match 6.3%; Score 253; DB 5; Length 1979;
Best Local Similarity 19.6%; Pred. No. 0.0083;
Matches 158; Conservative 163; Mismatches 253; Indels 234; Gaps 38;

QY 42 MHVKEKKNNDENKRDDEKRNKTOEBHLKEIMKIIVIEVKGGEAVYKKEAEKLEKVP 101
Db 329 LNKQEKEREKEREKEREKEREKEDYDLK-----ELKDEKI-----SILEKVS 376
QY 102 DYLEMYAIGKIYIVGDITK--HISLEALSEDKKKIKIDYGDALLHEHYVAKGEYE 159
Db 377 -----IVREMDIEKREHNFH--MEDQKOLKNSFYANN--NOLAKYKCE--- 418
QY 160 PVLYQSSSEYVENTERKALNAYVEIGKILSRDILSKINOPYQKFLDVLNTIRKASDSDQ 219
Db 419 -----IKMLTFELEKEKELK---DIENVSKKEIKLNQLENEKQKQIAFNKKNKE--- 468
QY 220 DLFTNLQKLEPDSVSEPLEQNSNEVQEVFAKFAFYIEPQHRDVLQVAPFAFNM 279
Db 469 -----IHGLKEELKE-SVKIRKIEFQELQEW-----VDIKQKEDQL--- 507
QY 280 FNEQINLSLEELKDQMLSRHY-----EKMEKIKOHYQHSMSLSSE----- 321
Db 508 YMQOIESISIFLSKKEKEKYQKNTYIEBINNLNEKLEFETKETNTQNNNTNINMLN 567
QY 322 ---GRGLKLLQPIPEKKDDI-----IHSLSQEEKEL---LKRQIDSSDFLSTE 368
Db 568 DIHMLNNGIKTMNQISTLKNDVHLNEQIDKLNNKGTLSKISELVQIMDL--KEEK 625
QY 369 EFLKLLQIDIRDSLESEKELNRIQVDSNPLSEKEREFLKAKL----- 414
Db 626 DFLNQIVDLSNQI---DLITRKMEKEKMKLEQENKRYKQEMELLRGNKSSENILNN 680
QY 415 DIQPYDINQRL-----ODTGLDPSJNDVRYQKRDIONIDALLHOS 459
Db 681 DEVCQDLKRLKISLSEKMKMKKEHDKKLAELKQDCOVR--IREMNEKNEKIMLKEE- 737
QY 460 TGSFLYKLYLYENMNTNNTLTVLGLADVSTQNTK--INRGIFNEEKNRKYSSIS--- 514
Db 738 -----YED-KTN---TLQEQNEDEKINTLKEQNEDEKINTLKEVEHKINTWKEE 781
QY 515 -NYMIVDINERPALD---NRLKWRIQLSPTPRAGL-----ENCK 551
Db 782 YEHKINTLNQONEHKINTLNQONEHKINTWKEEYEDKMNLTLEQNEDEKMSLKEEYENKI 841
QY 552 LILQNRIGLEIKDY--QIKQSEKEYIRIDAKVPRKSIDPTKIDQAOALNINO----- 601
Db 842 NQINSNNEIKKDYVNEYIEEVDKLKTYLDEK---KKQFKEIYAIKRAKHEKQILLTE 898
QY 602 -EWNKALGPRKYTKLITFNVHNRYASNIYVSAYLILN-EKMKNNIQS--DLIKRVTVLV 658
Db 899 MEELKQCRQDKYSGL-----YEKTI-KLISICQNTINIECCDDIENDDIIRIREYIINN 952
QY 659 NGRVFVFTDILPNIAEQYTHODEIYEOVHSGLYVPESRSILLHGPKSVYELNDSGFI 718
```


Qy	129	ALSDKKKKID	-----	VGK	-----	DALLNHN	-----	YAAEGEPRVLVIOSSDDYVENEK	176
Db	1803	LDVWEIKELKEVQAOTERDKLNKKVDEL	-----	KNNYLNINKEQEGHRL	-----	DYIK	-----	EK	1851
Qy	177	ALANYVEIGKILSDILS	-----	KINOPYOK	-----	FLVLVNTIKNASDGDGLLF	223		
Db	1652	IFENLYIKCSE	-----	QONNIDSDYNTIYAKQYIKLIEDVFKLSDLSMTIEKKNSVANLEIC	1909				
Qy	224	TNQ	-----	LKEH	-----	PRDVSVEFLQNSNEVOEVPKAPAYATIEPOH	262		
Db	1910	TNKEDIKNLKLHVKILANFESGIIYMSDTMTETPREPDELNDLL	-----	NLOLVEERKHETS	1967				
Qy	263	-----	-----	ROYLOL	-----	TAPAFYMKQKFN	282		
Db	1368	TLENDSOLELDHGSNDESIDNLKAYNYDIELHTYSTQILKYLNNIOKLGDCNDLVDK	2027						
Qy	283	-----	QEINLSLEELKQDQRYMSREYKEWKIKQNHQHSDSLSEEGRLKLKLDIPIEPK	337					
Db	2028	CKEIRELSTALYDKLD	-----	ITSVINRENDSDNNIDVSNKLN	-----	IDALDYNEKEKE	2081		
Qy	338	-----	-----	DIHLSLSEKELLKRIQIDSS	-----	DELSTEKEPLKLQIDIRSL	382		
Db	2082	IFDVAEEYKTLDDPTKKNAYKKAELIKLVNDINKTKEDLDIYENDDELEKSL	-----	TL	213535				
Qy	383	SEEKELNLRIVODSSNPLSKEKEPLKLKLDIQDYINORLODPTGGLSDPSINLDVR	442						
Db	2136	SSNMEIKTIYQ	-----	NSIYSFS	-----	DLINKIN	-----	DID	2164
Qy	443	KQYKRDIONDALLHQ	-----	SIGSTLYNKKIYLYENMMINNLATLADL	-----	VDSTDN	-----	493	
Db	2165	KEMKTLIPMLDELLNEGHNIDISLVNFI	-----	IRNIQIKGIDINKIRQENDNIC	2217				
Qy	494	-----	-----	TKINRGTFNEKKNFKYS	-----	ISANTVIYD	-----	INRPALDNR	531
Db	2218	FEYQNNYFNFKISDISEFNKYDHDIVKYNISNNIDVYKKNHSLSEHYATNTIEN	I	2276					
Qy	532	LKARIQLSPTPRAGYVLENGK	-----	LLIQRNIGLEIKQVQIKQSEKEYIRIDAKVYPRKSI	588				
Db	2277	MTSIVEINEDPTBMSLEEQDKLLELYEN	-----	KKEKNIINNKKIYHFNKL	2325				
Qy	589	DTKIQ	-----	EAQLNINDEWAKALCPKTYTKLTFPNVHNRYASVIVESAYILLNEMKNNIQS	645				
Db	2326	KEIENSLEIETYSISTFNFKI	-----	NETONNIDILKNEF	NNIKT	23633			
Qy	646	DLIKVYNYLVDOGGRVFPDITL	-----	PINARQYHODEIYEQVHVK	689				
Db	2364	KINDKVK	-----	ELVHVSPTLLESIQTENNLXGDIAMSNIDQYKREDINVELKV	2415				
Qy	690	GLVYPEERSILLHPSKGVGLRLRSDSEGFIHEFGVHADVAG	-----	YL	LDKMOQSLV	741			
Db	2416	KLYIENTITNLL	-----	GRINTFKELDKQODENNIGDKIETLEKNNSYI	2460				
Qy	742	TNSKKFLIDIFKEGSGNLTSGRTNEAEFF	770						
Db	2461	IKLKEKANLNKENFSKLLQIKKNETELY	2489						
RESULT	8								
Q9BK45		PRELIMINARY;		PRT;	3254	AA.			
AC	Q9BK45								
DT	01-JUN-2001	(Tremblrel. 17, Created)							
DT	01-JUN-2001	(Tremblrel. 17, Last sequence update)							
DT	01-JUN-2001	(Tremblrel. 17, Last annotation update)							
DE	REPLICOCYTE BINDING PROTEIN 2	HOMOLOG B.							
OS	Plasmodium falciparum.								
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.								
OX	NCBI_TaxID=5833;								
RP	SEQUENCE FROM N.A.								
FX	MEDLINE=21101060;	PubMed=11160005;							
RA	Triglia T., Thompson J., Cartuana S.R., Delorenzi M., Speed T.,								

RA Cowman A.F.:
RT "Identification of Proteins from Plasmodium falciparum That Are
RT Homologous to Reticulocyte Binding Proteins in Plasmodium vivax".
RL Infect. Immun. 69:1084-1092(2001).
DR EMBL: AF312917; AKK19245.1; -
SQ SEQUENCE 3254 AA; 362876 MW; 6F9CAFA5AA6167BA CRC64;

Query Match 6.2%; Score 248.5; DB 5; Length 3254;
Best Local Similarity 19.8%; Pred. No. 0.023;
Matches 184; Conservative 146; Mismatches 274; Indels 325; Gaps 43;

Qy	44	VKEKKKKDEKKRRDEERNKTOEHLKELMKHVLTEVGEAAVKEAAKL-----	95
Db	1684	LKEKONNEAEYKKNEMNYVNETE--KEITIKHKKNYEIRIMHKEITNEKKKKFMESSN	1742
Qy	96	-----LKEVSDVLEMYKAIGKRYIVDDITKHSILE	128
Db	1743	KSJLTLDSDRSRMEYNEXYINDYININENFEKHONILNEIYNGNESYNIINKRMETIIND	1802
Qy	129	ALSDEKKKIKDI---YGR-----DALLHEHY---VYAKEGEPLVYIOSSSEDEVENTER	176
Db	1803	LDYNEIKKEIKEVATQETDKLNKKVDEL--KNYLNINKEQEGHRLT-----DYIK--EK	1851
Qy	177	ALANYEYELGKLSNDLS-----KINOYQK-----FLDVLTINMADSODGODLLE	223
Db	1852	IFNLYIKGSE--QONIIIDSYNTYTKKOYIKTIEDYKFLDLSLTWTEKKMSVANLEIC	1909
Qy	224	TNQ-----LKEH-----PYDFSVEFLEQNSNEVCEVAFKAAVAYIEPOH----	262
Db	1910	TNRKDKINDLKKHYKLANFGIIVMSDITMETIPEPPLDNDLL--NIQLYFERKHETTS	1967
Qy	263	-----RDVLOT--YAPAFNYMDKENE-----	282
Db	1968	TLENDSDLELDHGSNDESIDMLKAYNDIIEHLYSTOILKYLLDNIORLKGDCNDLKYD	2027
Qy	283	-----QELNLSLEELKODRMLSREYKEKMEKIKOHQHMOSDLSREGGLKLTQIPIEPK-	337
Db	2028	CKEIRELSTALYDIKID--ITSYVNRENDISNNIDVSNKLN-----IDAIQYNEKYKE	2081
Qy	338	-----DIIHLSOEEKELLRKRIODISS-----DFLSTEEKEFLKLOIDIRSL	382
Db	2082	IFDWEVEKYTLDDTKNAYIYKKAELIKANDINKTKEDDIDYFNDDLEKLS-----TL	2135
Qy	383	SEEEKELLRIOVDSNPJSEKEKEFLKKLKDIDOPYDINOQLQDGTGLDLSPSINDVR	442
Db	2136	SSNMEIKTIVQ--NSYNSFS-----DINKNIN-----DID	2164
Qy	443	KOYRDIQNDALLHO--SIGSTLYAKITYLEEMNINNLATLGLD--VBDTN-----	493
Db	2165	KEMTLLIPMDELLEBENHNIDISLAFI-----IRNOIQIGINDIKNIREOEENDNIC	2217
Qy	494	-----TKINRGIFNEFKKNFYS--ISSNWIVD-----INERPALDNER	531
Db	2218	FEYIOWNYNEFKSDISIFENKYDHIKIVDNYISNIDVYKKNHSLSEHYVNTYNIEN-I	2276
Qy	532	LKMRIOLEPTRYAGYLENGK--LIIQRIIGLEIKDVOQIIOKSEKEYIRIDAKVYPSKI	588
Db	2277	MTSIVELNEDDEBMSLEEFODKLELEYENF-----KKEKNIIINNNKYKIVHFKL	2325
Qy	589	DTKIQ-----EAOULINDEWKNALGCPKYTKLITFNWNNRYASVIVESAULLNEMKNNIOS	645
Db	2326	KEIENSLFETVNSISTYFNKI-----NETONIDILKNEF--NNIKT	2363
Qy	646	DLIKKYVNYLVGDGRVFFDITL-----PRIAOLYHODEIYEQVYK	689
Db	2364	KINDVK-----ELVHVSFTLLESIQTFENNLVGDMSNIQDYVYKEDINVELKV	2415
Qy	690	GLVYEPSRSIILHOPSKGVELRDNDSGFIIEFGHAVDYAG-----YL-LDKNOSDLY	741
Db	2416	KLYIENTITNL-----GRINTFKELDKIQDENNGIDKYTEIKENKNSYI	2460
Qy	742	TNSKRFIDIRKEGSNLTSYGRTNEAEFF	770

DB 2461 IKLKEKANLKENFSKLLQNIKRNTEELY 2489

RESULT 9

09G276 PRELIMINARY: PRT: 1387 AA.

ID 09G276

AC 09G276:

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE RETICULOCYTE-BINDING PROTEIN 2 HOMOLOG A (FRAGMENT).

OS Plasmodium falciparum.

OC Eukaryote; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.

NCBI_TaxID=5633;

RM [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20402589; PubMed=10920203;

RA Rayner J.C., Galinski M.R., Ingravallo P., Barnwell J.W.:

RT "Two Plasmodium falciparum genes express merozoite proteins that are related to Plasmodium vivax and Plasmodium yoelii adhesive proteins involved in host cell selection and invasion."

RL Proc. Natl. Acad. Sci. U.S.A. 97:9648-9653(2000).

DR EMBL; AF196347; AAF98066.1; -

DR NON_TER 1

FT NON_TER 1

SO SEQUENCE 1387 AA; 164871 MW; 5660154744F0C285 CRC64;

Query Match 6.2%; Score 245.5; DB 5; Length 1387;

Best Local Similarity 19.9%; Pred. No. 0.011;

Matches 186; Conservative 143; Mismatches 269; Indels 337; Gaps 43;

DB 44 VKEKKNDENKRRKDEERNKQOEHLKIMKIVKIEVKEBAVKEAEKL----- 95

DB 48 LKEKONMEAYKKMEKNYVETE-KEIKKKKVEIRIMEHIKKEKKKKKESNN 106

QY 96 -----LEKPSVLEWKAIGKRIYDGIKISLE 128

DB 107 KSLTLMDSFMSFYNEYINDYNINENEFKHONILMEIYNGNESYNIINTMTETIINN 166

QY 129 ALSEDKKIKDI---YGR---DALLHEHY---VYAKEGEPLYVLOSSEDEYENFER 176

DB 167 LDYNEIKELKEVAQTEYDKLNKKVDEL--KNYLNKIKKEGHRLL-----DYIK--EK 215

QY 177 ALNVEYEGIKLSRDL-----KINOPYK-----FLDYLVNTIKASDS----- 216

DB 216 IFNLTKGSE--QONIIDDSYNYIVKQYIKTIEDVKFLDLSLNTIEKKNSVANLEIC 273

QY 217 -----DGGDL-----FTNOL 227

DB 274 TSKEIKILKLKIVILANFSGLIWSDTNETETTPENPLEDNDLNLQILFERKHEITSTL 333

QY 228 KEHPDPSVEFLQNSNEVEVFAKAFAYIIEPOHRDVLQLYAPEAFNYMDKFNE----- 282

DB 334 -ENDSDLELDHGSNSDESIDLN-KVYNDIE-----LHYSTQILKLDNIQKLGDC 385

QY 283 -----QEINLSLEPLKQRMKSRYEKWEKIKOHYQHSDSISEGRLKLGQIP 332

DB 386 NDLVDCCKELRELSTALYDLKIO-ITSVINRENDISNNIDIVSNKLN-----IDAIOVN 439

QY 333 IEPK-----DDIHSLSOEKELKRIQDSS-----DELSFEKEFLKLDI 376

DB 440 FEKYEIDDNVEEYKTLDDTKNAYIVKKAELIKNDINKTKEDLDLYFNDLDELEKSL-- 497

QY 377 DIRDSLSEBEKELLRKIQVDSNPLSEKEKEFLKLDIOPDYDINQRLQDTGGLIDSPS 436

DB 498 -----TSSNEMEIKTIYQ-NSYNSFS-----DINKNN----- 525

QY 437 INLDVARKQYKRIQNIIDALLHQ--SIGSTLYNKIYIYENMMNINLTATLGADL--VDSFD 492

DB 536 ---DIDKEKKTILPMDELLENGHNIDISLVNFI-----IRNIOIKIGNDIKNIREOE 575

QY 493 N-----TKINRGINEKKKFKPYS--ISSNYMIVD-----INERP 525

DB 576 NDTNICEFYIÖNNYNEFIKSDISIFNKYDHIKVDHIVYNNIDVYNNKNSLSLSEHYNAFN 635

QY 526 ALDNERLKWRIOISPDPTRAGYLENGK---LIÖRNIGLEIKDQVILKÖSEKEYIRIDAKV 582

DB 636 ILEN-IMTSIVELNEDETENSLEETQDKLEYLEN-----KKEKINNINNYKI 683

QY 583 VKRSKIDTKIQ---EQÖLINÖEMNKALGLPYTKLITFNHNRASNIVESAYLILNEM 639

DB 684 VFHNKL-KEIENSLEFYNSISTFNKI-----NETONIDILKNEF 722

QY 640 KNNIOSDLKKYTNLYVDNGRFVFTDIL-----PNTAEQYTHODELY 683

DB 723 -NNIKTKINDKVK-----ELVHVDSTLTLESIOTFNNLGDLSNIDQYAKIEDINN 773

QY 684 EQVHSGKLYVPESRSILLHGSPKGYELRDSGCFIEHFCHAVDDYAG-----YL-LDK 735

DB 774 VELKKVKKLYENTNLT-----GRINTFIKELDKYQDENNGIDKYEINK 818

QY 736 NOSDLVTSKRFIDITFEKGSNLTSGRTNEAEFF 770

DB 819 ENNSYIILKKEKANNLKENFSKLLQNIKRNTEELY 853

RESULT 10

QYVT6 PRELIMINARY: PRT: 1127 AA.

ID 09YVT6

AC 09YVT6:

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE ORF MSV156 HYPOTHETICAL PROTEIN.

GN MSV156.

OS Melanoplus sanguinipes entomopoxvirus (MeSPV).

OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;

OC Entomopoxvirus B.

OX NCBI_TaxID=83191;

RM [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TUCSON;

RX MEDLINE=99102612; PubMed=9847359;

RA Afonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.:

RT "The genome of Melanoplus sanguinipes entomopoxvirus.";

RL J. Virol. 73:533-552(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=TUCSON;

RA Afonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.:

RL Submitted (May-1998) to the EMBL/Genbank/DBJ databases.

DR EMBL; AF063866; AAC97677.1; -

SO SEQUENCE 1127 AA; 134265 MW; F185DAID5A3FE7D1 CRC64;

Query Match 6.0%; Score 239; DB 12; Length 1127;

Best Local Similarity 20.8%; Pred. No. 0.016;

Matches 159; Conservative 121; Mismatches 252; Indels 234; Gaps 35;

QY 60 ERNKQOE---HLKETMKHIVKIEVGEBAVKEAEKLEKPSVLEWKAIGKRIY- 115

DB 24 ENNKVSLDIINSLEYELNNI-----KFSDKTINFIKNNYKIVKIEKIFY 66

QY 116 -----IYGDITKHISLSEDKKIID-----YKGDALLHEHYVAKSEYEV 161

DB 67 MHNQFINDYNIILQYLIIFYENNEIKENKFNPCKNPLY-----NITYKKKLYIAD 118

QY 162 LYIÖSSEDEVMEKALNYYE---IGKILSDILSKINOPY-----QKLDVL 207

DB 119 L-----DIEKKDKELVYNIEOKNAVDKI--NDIKNNVNIHSDNETITTKETLIDIL 170

QY 208 NTIKNASDSGDGDLFTNOLKEHPDPSVEFLQNSNEVEVFAKAFAYIIEPOHRDVLQ 267

DB 171 NKLKLVSSDEKQLI--EQIYKNNKIEIF--KNIDVQKEINK----- 211

```
QY 268 LYAEAFNVDKREFOEINLSLEELKQDRMLSRKEWEKIKOHVQHSDSLSEGRGLK 327
      : : : : :
Db 212 -----KQDELNLKLDL-----SKKEFK----- 229
QY 328 KLDPIEPKKDDIHSOSEKELKRIQIDSSDFLSTEEKEFLKLDIROSLSPEEK 387
      : : : : :
Db 230 -----KQELKKTIDKQOEELIKLNDKEINFINIDEKQKLDQINSKI-NLNLNINIK 280
QY 388 ELLNRIOVDSNPLSEKEKEFLKLDIQPYDINQRIODTGGILDSPSINLDVRKOYKR 447
      : : : : :
Db 281 GVMN-LYTERKNKISNQNLEINLK-DSTIKSLDEKQKLDL-----ELDK 322
QY 448 DIONIDALLHQSISTYKNTIYLENNNINLRATLGCADLVDSDTNTRKINRGJFNEKKN 507
      : : : : :
Db 323 NINNI-----TSLYKN-----SNRKINIQDLLESSLTD-FNNANIN-----INELKSK 365
QY 508 FKYSISNMYIVDINEBPALDNERLKWRIQLSPDTRAGYLENGKLLIQRNIGLEIKVQI 567
      : : : : :
Db 366 IK-----LFD-NDQKLANNDTEONNKIT-----DFNNSTRIRKEKLDTEYKKIDD 411
QY 568 IKOSEKEYIRIDAKVYPKSKIDTRIQEAOQINQWENKALGLPKYTKLITFNVHNR----- 623
      : : : : :
Db 412 IKNNNLOKLEESYK-----KIDEOTEYKKNKINKEYNDIEL-KNNMLQKLEENKKIDE 465
QY 624 ---YASNIVESAVLILNEMK-NNIQ-----SDLIKVTNVLVDGNGRFVPTDITLP 670
      : : : : :
Db 466 QTEYKKNKIKETNDIETLNNNNLOKLEENKKNINDKLTILKNDIESNTLFI-----NKL 520
QY 671 NIAEOYTHODEI-----YEOVHSKGLYVPESSKILHGFSGVLENDSEGFIERGHGA 724
      : : : : :
Db 521 NIDFKKSKREIKLNTEYQOL-----RKDLLENINKTLEMLKLDLNDKLSLEQL 570
QY 725 VDDYAGYL--LDKNQSDVLVNSKKFLDIFKEEGSNLTSYGRTEAE 768
      : : : : :
Db 571 YDSKKNILDQIDKIYNSLKEKNKIDEXF---SNIEKFIYVIE 612

RESULT 11
Q9EMP3 PRELIMINARY; PRT; 1238 AA.
AC Q9EMP3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE AMV156.
OS Amsacta moorei entomopoxvirus (AmEPV).
OC Viruses: dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OX NCBI_TaxID=28321;
RN RN
RN SEQUENCE FROM N.A.
RX MEDLINE=20396580; PubMed=10936094;
RA Bawden A.L., Glassberg K.J., Digans J., Shaw R., Farmerie W.,
RA Moyer R.W.;
RT *Complete Genomic Sequence of the Amsacta moorei Entomopoxvirus:
RT Analysis and Comparison With Other Poxviruses.*;
RL Virology 274:120-139(2000).
RN RN
RP SEQUENCE FROM N.A.
RA Bawden A.L., Glassberg K.J., Digans J., Shaw R., Farmerie W.,
RA Moyer R.W.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF250284; AAG02862.1;
SQ SEQUENCE 1238 AA; 147096 MW; 9BBF80A39DB6E35B CRC64;
```

```
Query Match 5.98; Score 237; DB 12; Length 1238;
Best Local Similarity 20.1%; Pred. No. 0.022;
Matches 189; Conservative 170; Mismatches 293; Indels 290; Gaps 52;
QY 56 RKDEERKKTQDEHKLKMKIIVKIEVGGEAVKKAELKLEKVPDVLKRYKAIGKI- 114
```

```
Db 158 RSODEFINIKNIYHDIKN-----DNNIIDQLISRPIDITILKNEFLDKLT 205
      : : : : :
QY 115 -----YVDDIDITRKHS-----LEALSDEKKIKINDIYGKALLHEHYV-----AKEG 157
      : : : : :
Db 206 NYIKNIYISDDKLVNIISKNPFLNNIKIYENINIKINNIILNK-IYVMDPDLFIYNNVKN 264
QY 158 YEPVLVQSSEDEVVENTEKALNYYEIGTLISRD-----ILSK-----INQYOKF 203
      : : : : :
Db 265 YDFINLFEKNNEDEFIPLKNIINNDY-INNLISSEYKNIIDMEIKSKNIDIIMDYEVKN 323
QY 204 LDVINTIKNASDSGQ-QDILFTNQ--LKEHPIDFSVEFLEONS-----NEVQ 247
      : : : : :
Db 324 SDVIYKLLQKISDEYVSNLISLENFLKIKDIISSENIDISLNSLNIENKFKILK 383
QY 248 EVPAKAFAYYIEPOHRDVLQ-----YAEAFNVMKFN--EOEINLSLEELK--- 293
      : : : : :
Db 384 EIINKEL-----KGNIDMLVNNFNNDYKQYKKSIDEKIDIDINIKODEINTHLSLDVLI 438
QY 294 -----DORMLSRYEKWEKIKOHVQHSDSLSEGRGLKLT 329
      : : : : :
Db 439 SPYONLRTYANNIDITMESIYDYKQILNLVQETEKLDHYKQWVNSRFRQ-----LSDV 494
QY 330 QIPLEPKKD-----IHSLSOE-----KELKRIQIDSSDFLSTEEKEFLKIQ 375
      : : : : :
Db 495 SNSIDEKYDNFKNQSAIILKINDKESFYNYVEDISKIKLINDDL-IFNNIIFENLN 553
QY 376 -----IDIRD-SLSEEEKELNRIOVDSNPLSEKEKEFLK-----KLKLDIQPYD 420
      : : : : :
Db 554 SGDVNTNIDIKNIDLSEDIKIFINKI-----LPELIKFNGLGEVYNSDKFNIINESV 606
QY 421 INQRLDQDGLDPSINLD-VKQYKRDQNDALLHQSISTGLYV-----KIYLENMN 475
      : : : : :
Db 607 IN-----NINTDYILNIKISNIPTESSII-TVEESRLKXYIDDKFY-----D 647
QY 476 INNLATLQADLVDSPTNTKINGIFNEFKNFKYSISSYVW---IYDINERPALDNERL 532
      : : : : :
Db 648 INNSISLINTDI--STNNAKISE-IAADIER-KNNISENIKTIPLDFDEKAKIDKIL 703
QY 533 KWRIOISP-----DTRAG-YLEN-----GKLLIORNIGLE----- 561
      : : : : :
Db 704 KY-LEINPVLQSKYANIKIKEYISYTSRASLOQIKNIITNLIDKILSDNKLINIELSNFDS 762
QY 562 -----IKDVQIIKQSEKEYIR---IDAKVYPKSKIDTRK---IQEAOQINQWENKAL-GL 609
      : : : : :
Db 763 LDDKILKSIDLKLNDQDEIKELLSKRTETISEFEDEKINAINELLLPLESKNENIAGI 822
QY 610 PKYTKLITFNVHNRVYASNIVESAVLILNEMKNNIQSDL-----IKKYTNVYLVDGNGRF 662
      : : : : :
Db 823 YDEIINIIS-KNYNEILFKIDNNMYLOLEEKIENVLDNINVDITIKNNIKKEIND-F 880
QY 663 VFTDITLPNIAEOYTHODEI---YEOVHSKGLY-----VPESRSILHGP 704
      : : : : :
Db 881 IYKNYSOGNIPKYITERENTLRCGAETLKKIITYINSLNENAPRINFESVPKT-NVSKSK 939
QY 705 SKGVELRN-----DSEGFIEHFGHAV-----DDYAGLILDK----- 735
      : : : : :
Db 940 STNDVANLIIKYNMKEKKNPRLQOAGFVOEYKPIPLINVIYDPLNKYINKINTENIKI 999
QY 736 -NOS-----DLVYNSKKFIDIFEKESNLTSGYRTNEAEFF 770
      : : : : :
Db 1000 LNNSYKKILFEYDNNNSINDIIR---NIITY-KTEDIISY 1036

RESULT 12
Q26023 PRELIMINARY; PRT; 1048 AA.
AC Q26023;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOHETICAL PROTEIN (FRAGMENT).
OS Plasmodium falciparum (isolate 3D7).
```

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RA Alano P., Elliott J.F.;
RL Submitted (Apr-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; M69147; AAA74653.1; -
KM Hypothetical protein.
FT NON_TER 1
FT NON_TER 1048
SQ SEQUENCE 1048 AA; 126518 MW; 7AF051480FA22424 CRC64;

Query Match 5.9%; Score 235; DB 5; Length 1048;
Best Local Similarity 19.8%; Pred. No. 0.022;
Matches 174; Conservative 149; Mismatches 301; Indels 254; Gaps 43;

QY 27 FIPVQAGAGHD-----VGMHVKEKKNKDKENKRDENKTOBEHLKEIMKHIVKIE 80
DB 295 FTTISGEENHONRELKRIEANKIEEMKKRFNEQOEQRKKKAE--DENMETIOKHD 352
QY 81 VKGEAVKKEAAKLEKVPDLEMYKAGKITYVDG---DIRKHISLEALSD--K 134
DB 353 METSKLEKKEEDEVTODEEFD-----KQYAIIDQEELEFLRTDSEGESDVPK 402
QY 135 KKIK-----DIYKDALHEHYAAKEGYEPVUVIOSSEDY-ENTEKALNYEYI- 184
DB 403 DKVKPRPDGSPSPFYNNAISSFHEKMEELYN--TSISSLNVYKEINKRPDQVYKELK 460
QY 185 -----GKILSRDIISKINO-----PYQ-----KFLDVLNTIKNASDSG 218
DB 461 SKTYPRFDLTSQTKNCNKLFGKLNETIKDKEYOKNIOSYKKNVIDIIDIO--KKANG 518
QY 219 QDLFTNOLKEHPTDSVEFLDONSNEVOEVPKAF----- 254
DB 519 KYIIIONLIEIKIDYKGVNRLSDRKFYKNFKVLGKKRKKMLLEDFFRAQFGAIRFIKD 578
QY 255 -AYIEPOHROVLO-LYAPFAFNMDKFNEDQINLSBELDKORMLSR--YEKMEKIK 308
DB 579 LTTTYMEETKYVEDIYMEKKYKKEYS-----KMRRIISSMLDYEVMKQIK 627
QY 309 OHYOHMSDSLSEGRGLKLLKQIPIEPKKDDIHSLSOEKELKRIQIDSSDFLSTEER 368
DB 628 EHY-HKVDITISEH-----KFOEIRQHMROKIENTIHELKEMKYVQIQDITNYY----- 675
QY 369 EFLKKIQIDIRSLSEEEKELNRIQOVSSNP-----LSEKKEFLKKLKDIOPIYIN-- 422
DB 676 -----HOLENINISELLQALQONKNIPRHILNVLEKLEITRKRRKN-KP-DISTS 722
QY 423 -----ORLODTGGLIDSPSINLDVARKOYKRDIONDALLHOSIGSTLYNKIYLENMN- 475
DB 723 SHATDQOVSDT--LIRGAHNHGDIIKGEDND---EVLLIEIOQL--KTKMGONQNO 773
QY 476 -----INNLATATGADLVDS---TDNTKINRGIFNEFKKFKYSISSNYIVADINER 524
DB 774 VGSILEKLNLS-----DQYQLQDKLVNVEDIYKNL-RNFKIYIEKLHESKINNE 824
QY 525 PALDNERLKWRIQSPDTRAGYLENGKLLIQRNIGLEIKDVO-----TIKOS 571
DB 825 KFI-----TKVDVLSNVYSTLEEVYKFKLLDFQEMSPSEKDELKHELKYLEE 870
QY 572 EKEYIRIDAKKVPKSKIDKIKQEAQ--LNIQEMKAKGALPKRYTKYLTFNNVHNRASNI 628
DB 871 RKKYITLLEIQI--RDTLSNIONGEGDHTINNNNNN--NANNNLKKOY 914
QY 629 VSAVALILNEMKNNI--OSDLIRKVT---NYLVDSNGRFEVTDITLPNIAEQVTHODEYI 683
DB 915 LKDLDLEISKLGHILEVDIKKNIALEQINYLNN-----TNEYIPDVIKRLMPAPRIY 968
QY 684 ---EOVHSGGLYVPSRSKILLHGPKSGVELRDNDSGFIHFGHADVAGYLLDKNQSDI 740
DB 969 PVSEDIYDITVWVNDNTAV-----INNTLRFVMTFQOKIYDYDHLI-----F 1012

QY 741 VTNSSKKFIDIEKE-----ESSNLSYGRTNAEFFAEAF 774
DB 1013 VYNIKEL--TYKENLADKEYNANYTEMNRFYHLHLEEF 1048

RESULT 13
ID 025857 PRELIMINARY; PRT; 3119 AA.
AC 025857;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PG377
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96360472; PubMed=8719156;
RA Handman E., Osborn A.H., Symons F., van Driel R., Cappel R.;
RT "COS cell expression cloning of Pfg377, a Plasmodium falciparum
RT gametocyte antigen associated with osmophilic bodies.";
RL Mol. Biochem. Parasitol. 74:143-156(1995).
DR EMBL; I04161; AAC37257.1; -
DR InterPro: IPR001064; CrystalIn.
DR PROSITE: PS00225; CRYSTALIN_BETAGAMMA; UNKNOWN.1.
SQ SEQUENCE 3119 AA; 377358 MW; 519B99D5BDFEFC CRC64;

Query Match 5.9%; Score 235; DB 5; Length 3119;
Best Local Similarity 19.8%; Pred. No. 0.082;
Matches 174; Conservative 149; Mismatches 301; Indels 254; Gaps 43;

QY 27 FIPVQAGAGHD-----VGMHVKEKKNKDKENKRDENKTOBEHLKEIMKHIVKIE 80
DB 964 FTTISGEENHONRELKRIEANKIEEMKKRFNEQOEQRKKKAE--DENMETIOKHD 1021
QY 81 VKGEAVKKEAAKLEKVPDLEMYKAGKITYVDG---DIRKHISLEALSD--K 134
DB 1022 METSKLEKKEEDEVTODEEFD-----KQYAIIDQEELEFLRTDSEGESDVPK 1071
QY 135 KKIK-----DIYKDALHEHYAAKEGYEPVUVIOSSEDY-ENTEKALNYEYI- 184
DB 1072 DKVKPRPDGSPSPFYNNAISSFHEKMEELYN--TSISSLNVYKEINKRPDQVYKELK 1129
QY 185 -----GKILSRDIISKINO-----PYQ-----KFLDVLNTIKNASDSG 218
DB 1130 SKTYPRFDLTSQTKNCNKLFGKLNETIKDKEYOKNIOSYKKNVIDIIDIO--KKANG 1187
QY 219 QDLFTNOLKEHPTDSVEFLDONSNEVOEVPKAF----- 254
DB 1188 KYIIIONLIEIKIDYKGVNRLSDRKFYKNFKVLGKKRKKMLLEDFFRAQFGAIRFIKD 1247
QY 255 -AYIEPOHROVLO-LYAPFAFNMDKFNEDQINLSBELDKORMLSR--YEKMEKIK 308
DB 1248 LTTTYMEETKYVEDIYMEKKYKKEYS-----KMRRIISSMLDYEVMKQIK 1296
QY 309 OHYOHMSDSLSEGRGLKLLKQIPIEPKKDDIHSLSOEKELKRIQIDSSDFLSTEER 368
DB 1297 EHY-HKVDITISEH-----KFOEIRQHMROKIENTIHELKEMKYVQIQDITNYY----- 1344
QY 369 EFLKKIQIDIRSLSEEEKELNRIQOVSSNP-----LSEKKEFLKKLKDIOPIYIN-- 422
DB 1345 -----HOLENINISELLQALQONKNIPRHILNVLEKLEITRKRRKN-KP-DISTS 1391
QY 423 -----ORLODTGGLIDSPSINLDVARKOYKRDIONDALLHOSIGSTLYNKIYLENMN- 475
DB 1392 SHATDQOVSDT--LIRGAHNHGDIIKGEDND---EVLLIEIOQL--KTKMGONQNO 1442
QY 476 -----INNLATATGADLVDS---TDNTKINRGIFNEFKKFKYSISSNYIVADINER 524
DB 1443 VGSILEKLNLS-----DQYQLQDKLVNVEDIYKNL-RNFKIYIEKLHESKINNE 1493

```
OY 525 PALNERRKMRIOISPTDRAGYLNGKLLIORNIGLEIKVQ-----IIKOS 571
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1494 KFI-----TKVDLSNVYSTLEYWKFLLHDQEWSEFEKELEKHLYLEE 1539
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 572 EKEYIRIDAKVVPKSKIDTKIOEAO-----LNIQEMNKALGLPKYTKLTTFVHNHRYASNI 628
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1540 RKKITTEIQI--KDTLSTNIQNGEGDHINNNNN-----NVRNNLKOV 1583
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 629 VESAYLILNEMKNNI--OSDLIKKYT-----NYLVDNGRFVETDITLPNIAEQYTHODEIY 683
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1584 LKLDLEISKLGHLILEVDIKKNIALDEQIVLYTNN-----TNETVDPVINDLMPAPRIY 1637
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 684 ---EQVNSKGLYVPESSILHGPSKGYELRNDESGFIHEGHAVDYAGIYLDKKNOSDL 740
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1638 PVSDIDYITWVRNDNTAV-----INNTRLHFVMTFDOKIYDDHLI-----F 1681
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 741 VTNSKKFIDIFKE---EGSNLTSGYRTNEAFFAEAF 774
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1682 VYNIKEI--IKENLADCKYNNANYTEKNNRYLHLEEF 1717
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
OY 025662 PRELIMINARY; PRT: 1939 AA.
AC 025662:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-AUG-1998 (TREMblrel. 07, Last annotation update)
DE REPEAT ORGANELLAR PROTEIN.
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5825;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=96V;
RA Werner E.B., Taylor W.R., Holder A.A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U4145; AAC63403.1;
SQ SEQUENCE 1939 AA; 229001 MW; B36E462001C6F22F CRC64;

Query Match 5.9%; Score 233.5; DB 5; Length 1939;
Best Local Similarity 19.4%; Pred. No. 0.053;
Matches 187; Conservative 170; Mismatches 280; Indels 325; Gaps 48;
```

```
OY 377 -----DIRDISEEKEELNRIQVSSNPLSEK-----EFLK-----KLKL 414
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1270 ILDSLNELINLENKKNVLTDENNNLKKEIEI--KKNKINKEKENNENTELINLNDIILKK 1328
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 415 DIQYDINORLODPGGILDSPSIL-----DYRKQYRKDIONIDALLHOSIG--STLYNK 467
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1329 EISEW-----KDEEKLTKENIKLNDIEQINEKYIKKEENLMIKFENENIYEVLSLKNQ 1382
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 468 IYL-----YE-----NNNI--NLATATGADLVSDTN--TKINGIFNE- 503
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1383 IEIKMKLEELNKKYELLAEKRETNMSISNDKKIYENNLLEPTDSKONNLKNKNEVDKT 1442
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 504 -----FKKNKYSISSNMYIVDINERPALDNERLKMRIQSPDRAGYLNGKLLIORNI 558
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1443 GDDINCEKNNDQAKESYIKDEIKKISMLYGEEL-----NKNKSYDEKVK-----NL 1489
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 559 GLEIKDQI-----IKOSEK-----EYIRID 579
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1490 TNELKELTKRNKGEBAIAELINKLNKEKNKSVKQNDSESSNNIITRKDGKTPREYVSD 1549
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 580 AKVVP--KSKIDTKIOEAOININQEMNKALGLPKYTKLTFFVHNHRYASNIVESAYILN 637
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1550 DKIOCKMKANVLTLEKKEP-----DLMDNINSLK-----ENFRVMSIVKE----- 1590
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 638 EKNNNIOSDLIKKYTNLYLVGNGR-----FVFPTDI--TLPNIAEQYTHODEIYQVH 687
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1591 --NNKVNQDKIVGIVSYFKCKCEKELKMDLVICLKDILSILPLNDNFVALFEKIDIL 1648
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 688 SKGLYVP--ESRSILHGPSKGYELRN-----DSEGF--HEFGHAV-----D 726
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1649 WKQWITPEITILFLRFSFLDKLRNYYKCVNEEYVNNERYISWALFQYLETASNLKK 1708
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 727 DYAGYLLDKKNOSDLVTNSK-----KFIDI-----FKEGSNLT 759
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1709 EMIVYLEKAEKDCENNSSFDPKPTDILNFSKDSIRLTKIAQLRKELNFEFEAKNIL 1768
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 760 SY 761
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1769 NY 1770
```

```
RESULT 15
OY 0900P0 PRELIMINARY; PRT: 1786 AA.
AC 0900P0:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE LIVER STAGE ANTIGEN-3 PRECURSOR.
GN LSA-3.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5839;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RA Daubersies P., Thomas A.W., Millet P., Brahimi-Zeghidour K.;
RT "LSA-3, a conserved pre-erythrocytic malaria antigen can induce
  protection in chimpanzees.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ007010; CAB55343.1;
DR InterPro: IPR001313; PUM.
KW Signal.
FT SIGNAL 1 63 POTENTIAL.
FT CHAIN 64 1786 POTENTIAL.
SQ SEQUENCE 1786 AA; 200101 MW; 5DF536D75B1BD98 CRC64;

Query Match 5.7%; Score 228.5; DB 5; Length 1786;
Best Local Similarity 18.9%; Pred. No. 0.078;
Matches 155; Conservative 134; Mismatches 271; Indels 259; Gaps 32;
```

QY	44	VKEEKKKDKNRKDERNKTOEHLEKIMKHVLEVGGEVAKAEKAEKLEVP----	100
Db	891	IEELEELEHENVUSALENTOSEEKEEVIDY-----EVEKEEATLLIEVEQAE	942
QY	101	-----SDVLEMAKIGAKIY-IDVG-DITYHISLEALSE----	132
Db	943	EKSANTITELFENLEENAVSNENVAENLEKLENTFNVLDKVEEYVEISGESLENEM	1002
QY	133	DKKIKIDYKDALHHEHYVAK-EGEYPLVIOOSE-----DYVENTEKAL	178
Db	1003	DKAFSEIFPNVAKGIEENLTGMSFISFIYIOSEEEKVDLENVVSILLINIEEMKGL	1062
QY	179	-----NYVEETKILSR-----ILSKINQYQKFLVDLNTIK	211
Db	1063	LAKLENISSIEGVOEYTEHVEBQNVYVDVPMKDQFLGILINEAGLKEMFENLEDFK	1122
QY	212	NASDSGDQLFTNOLKEHFTDPSVEFLBQNSNEVOEYAKAFAYIEPQHRDYLQALAP	271
Db	1123	SES-----DVIYVEEIKDEP-----YQKEVEKTSVITDEMEENIYDVAEE	1163
QY	272	EAFYIMKF-----NEQFINSLBELKDQKMLSNYEKWEKIK	308
Db	1164	EKEELDKMIDAVEESIEISDSKEETESIKQEKEDVSLVEYVDNDMDSEVEKVELTK	1223
QY	309	QHYQHW-----SDSLSEGRCL-----LKKLOIPIE-----PKKD	338
Db	1224	NMEELMKDAVELINDITTSKLIETQOLNEVADLIKDMEKLEKALSSEDSKEITIDAKD	1283
QY	339	DIHLSQOEKEL-----LKRIIDS-----SDFLSTEE-----KEF-----	370
Db	1284	DLEKYVLEEHDITTLDEVEVELKDVEEPIKEKVSOLKLEEDILKEVKEILESEILE	1343
QY	371	-----LKKLOIDIRDSLS-----EBEKEFLNRIOVDSSNPLSEKEPEKTKLKIDI	416
Db	1344	DYKELKTIETDLEBEKKELKDHFEKFEBAEBIKDLEADLIKVESSLEVEEKKLE---	1400
QY	417	QYVDINORLOD-----TGGLDPSINIDVQKQKRYDIONIDA	454
Db	1401	EVHELKEVEEHIIISGOAHIKGLEEDDLEVEDLKSGIIDLMLKGOMELDMOKESLEDTY	1460
QY	455	LHMOISOTLYUNKIYLEENMINNLTAITGADLVDSDTNFKINRG-----IFNEKKNF	508
Db	1461	KLGERVYSLK-----DVLSSAAGMBDEOMKTRKKAKORPKLEEVLLKEVEKEEP	1508
QY	509	KYSLSNMYMIVDINERPALD-----NEBKLMWRLO--LSPDTRAGLYENGKILLONI	558
Db	1509	KKKTKKKKVFODIKDEPRDEIYVEKMKDEIDIEVEDIEEDIEEDKVEDIEDDEDIEDI	1568
QY	559	GLEIKD--VOIKQSEKEYIRIDAKVVPKSKIDTKIOEAOLNINOEMWKAISLPRKYTLI	616
Db	1569	G-EDKDEVILDIQOKERIEKIVAK---KKKLEKVEEGVSGSLKKHNVDEM---KYVQKI	1621
QY	617	TFWYHNKYASNIYESAVLLINEMKKNIOGDLIKKATYNYL	655
Db	1622	DKEV-DKEVSALESKNDVTNNVLKON--ODFFSKKPNF	1657

Search completed: December 2, 2001, 13:54:16
Job time: 491.5sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:54:16 ; Search time 132.69 Seconds
(without alignments)
302.047 Million cell updates/sec

Title: US-09-747-521-2_COPY_9_282
Perfect score: 1404
Sequence: 1 KVISMSCVATITLISGPVFI.....RDVQLVAPFAFNWMDKENE 274

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147.5	10.5	1946	5	097291 plasmodium
2	135.5	9.7	1408	5	09NEU7 caenorhabd1
3	130.5	9.3	513	10	09W955 nicotiana t
4	129.5	9.2	980	5	096246 plasmodium
5	129	9.2	1245	5	096195 plasmodium
6	125.5	8.9	1510	5	025920 plasmodium
7	125	8.9	558	5	096224 plasmodium
8	124	8.8	649	5	018244 caenorhabd1
9	123.5	8.8	880	1	09WZC8 pyrococcus
10	122.5	8.7	508	5	096146 plasmodium
11	122	8.7	497	2	051503 caenorhabd1
12	122	8.7	1624	5	091121 borrelia bu
13	121.5	8.7	1661	5	006166 plasmodium
14	120.5	8.6	1156	2	096878 aquifex aeo
15	119	8.5	2269	5	026223 plasmodium
16	119	8.5	2747	5	09B3X3 plasmodium
17	118.5	8.4	440	5	097369 plasmodium
18	118.5	8.4	652	5	09U0D4 plasmodium
19	118	8.4	1005	1	058718 methanococ

20	118	8.4	2771	5	026216 plasmodium
21	117	8.3	1169	1	059037 methanococ
22	117	8.3	1351	5	096242 plasmodium
23	116.5	8.3	652	3	074974 schizosacch
24	116	8.3	999	5	0905A3 plasmodium
25	116	8.3	1087	5	096923 dictyostel1
26	116	8.3	1387	5	096276 plasmodium
27	116	8.3	1624	5	09U679 strongyloce
28	116	8.3	3130	5	09BK46 plasmodium
29	116	8.3	3254	5	09BK45 plasmodium
30	115.5	8.2	248	13	091866 xenopus lae
31	115.5	8.2	886	1	029230 archaeoglob
32	115	8.2	503	5	025777 plasmodium
33	114.5	8.2	1365	5	09NEV4 caenorhabd1
34	114.5	8.2	2274	5	09VYU0 drosophila
35	113.5	8.1	823	6	09TVY2 canis fam11
36	113.5	8.1	3351	5	094907 drosophila
37	113	8.0	435	2	0923C0 chlamydia p
38	113	8.0	593	2	044169 actinobacil
39	113	8.0	975	5	09V7I9 drosophila
40	113	8.0	1280	10	09SRD2 arabidopsis
41	113	8.0	1979	5	096133 plasmodium
42	113	8.0	2231	2	092KK7 helicobacte
43	112.5	8.0	532	11	09EQ77 mus musculu
44	112.5	8.0	653	2	034097 streptococ
45	112.5	8.0	823	11	P70451 mus musculu

ALIGNMENTS

```

RESULT 1
ID 097291 PRELIMINARY; PRT; 1946 AA.
AC 097291;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DD 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOHETICAL 231.8 KDA PROTEIN.
OS Plasmodium falciparum (Isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RA MEDLINE=99376085; Pubmed=10448855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Fellwell T.,
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum."
RL Nature 400:532-538(1999).
DR EMBL, AL034559; CAB39037.2;
DR InterPro; IPR002048; EF-hand.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_2.
KW Hypothetical protein.
SQ
SEQUENCE 1946 AA; 231792 MW; 59AC248AB980BE34 CRC64;

```

```

Query Match 10.58; Score 147.5; DB 5; Length 1946;
Best Local Similarity 25.58; Pred. No. 8.7;
Matches 69; Conservative 55; Mismatches 82; Indels 65; Gaps 15;

QY 31 DVGHNHVEKEK-----NKDENKRDDEERN--KTOEHNKKEIM-----KHIVK 70
DB 1360 DLKPTLKEKEKIQENYKNKINEHIKIIEELQSYKEIKKHKETIQIKEQJHNDKHAQ 1419
QY 71 IEVNG---EEAVKKKAEEKLEKVPDVLEMYKKAIGCKIYIVDGDITKHISLEALSEDPK 127

```

Db 1420 IOIEKNNEESYLKKNEN--EKMSNILEEKYKEL--STYEIDKNINK-IKIEDEKDE 1474
QY 128 KIKDIYGDALL--HEHYVAKGEYEPVLVIQSSSEYVENTEKALNYYEIGKILSRDL 185
Db 1475 NI-----LITKNEEINNLKEEYKAV-----OOHLEDI-----NVLVEKOKL-----AI 1512
QY 186 SKINOPYOKFLDVNTITKNASDSOGDLLFTNOLKEHPTDF--SVEFLQNSNEVQEVFA 243
Db 1513 DITTEKNNTINECDKTKNNK-----LNKKLENONNEHTLNTIKENQOITEREK 1566
QY 244 KAFAYIEPOHRDVLQLYAPAFNMDKFE 274
Db 1567 KNFTQKVES-----LEHAFKOSYNOLKDQNE 1592

RESULT 2
Q9NEU7 PRELIMINARY; PRT: 1408 AA.
AC Q9NEU7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Y39B68.M PROTEIN.
GN Y39B68.M.
OC Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Sulston J.E.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99065613; Pubmed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; ALI32896; CAB60918.1;
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; TYR_prot_phphatase.
DR Pfam; PF002206; MSN. 1.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRTYPPHPTASE.
DR SMART; SM00453; MSN; 1.
DR SMART; SM00453; MSN; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
SQ SEQUENCE 1408 AA; 159649 MW; AC8BDD4E56A4BAA CRC64;

Query Match 9.7%; Score 135.5; DB 5; Length 1408;
Best Local Similarity 26.9%; Pred. No. 25;
Matches 66; Conservative 34; Mismatches 74; Indels 71; Gaps 12;
QY 36 VKEKE--KKNDEKRRKQDEERKNTQOEHLKIMKHIYIEVKGEEAVKKAELKLEVP 92
Db 909 IKKEEYVAKAREKVKKEEYKAEERKKAEEERKKEK--VEKEKYAAEEKVKVEKERKKEVK 963
QY 93 SDVLEMYAIGKIYIVDGI--TKHISLEALSEDDKKIKDIYKALLHENVAKGEYE 151
Db 964 AEE--ERLAAEEKVKYIEKVKAEERKKAEEERKKEK--AEDA-----KVKEGOE 1013
QY 152 PVLV-----IQSEEDYVE-----NTEKALNYYEIGK-----ILSRDLISKINO 190
Db 1014 KAILINAAVQEKDEKSEARRKQKEMDNAAEKULNVEKKEKRIKLEKIMMSDY--KDDQ 1071
QY 191 PYQKFLDVNTITKNASDSOGDLLFTNOLKEHPTDFSVLEQNSNEVQEVFAKAFAYI 250
Db 1072 RTAFKCNNTIESVASTID-----KDHITE-----SYTYLP 1102

QY 251 EPOHR 255
Db 1103 EDKRR 1107

RESULT 3
Q9LM95 PRELIMINARY; PRT: 513 AA.
AC Q9LM95;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE KED.
GN KED.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots.
OC Asteridae; eusterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RX Pubmed=10945337;
RA Hara K., Yagi M., Koizumi N., Kusano T., Sano H.;
RT "Screening of wound-responsive genes identifies an immediate-early
expressed gene encoding a highly charged protein in mechanically
wounded tobacco plants."
RL Plant Cell Physiol. 41:684-691(2000).
DR EMBL; AB009883; BAA95789.1;
SQ SEQUENCE 513 AA; 61019 MW; P9598A37A840B970 CRC64;

Query Match 9.3%; Score 130.5; DB 10; Length 513;
Best Local Similarity 28.3%; Pred. No. 14;
Matches 56; Conservative 29; Mismatches 68; Indels 45; Gaps 9;

QY 37 KEKEKKNDEKRRKQDEERKNTQOEHLKIMKHIYIEVKGEEAVKKAELKLEVP 95
Db 314 KKKKKNKKKKEKTKDEKRYKSESEEDKKEETEEDKDDGQKKE--KEKKKKKKK 372
QY 96 LEMYAIGKIYIVDGI-----TKHISLEALSEDDKKIKIND-----IYK 135
Db 373 KDKKKVKSKESEDEDDQKDYNEVEVATREIKI-----EDDKKISDGADEKGRKKEGK 428
QY 136 DALHEHYVAKGEYEPVLVIQSSSEYVENTEKALNYYEIGKILSRDLISKINP----- 191
Db 429 D-----SKDEKQK-----DAKKDKAEKTRKLEDDKYSNGKLSK--LEKINAKLEAL 473
QY 192 YOKFLDVNTITKNASDS 209
Db 474 QOKKADIMKTIKAEEDKN 491

Query Match 4
Q96246 PRELIMINARY; PRT: 980 AA.
AC Q96246;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HYPOTHEETICAL 118.9 KDA PROTEIN.
GN PEB0765W.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=99021743; Pubmed=9804551;
RA Gardner M.J., Tetteelin H., Carucci D.J., Cummings L.M., Aravind L.,
Koonin E.V., Shellen S., Mason T., Yu K., Fujii C., Pederson J.,
Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Partea M.,
Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;

```

0Y      1 KVISMSLVAITLTSQVFLPVLQAGAGHSDVGNHYNEKSN--KIDNNKKDEERKNTQE 58
      1 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      456 KLAEMCKNHSIFNEKHYI-----SEGDKKIILFDKERNVKEDEDEDDYDINDND 508
      1 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
0Y      59 EHLKEIMKHIVKIEVGEAAVKKAEAKLLEKVPDYLEWYKALGCKIYLVDDITRKHS 118
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT	7	
ID	096224	PRELIMINARY; PRT; 558 AA
AC	096224;	
DT	01-MAY-1999	(TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE HYPOTHEITICAL 66.2 KDA PROTEIN.
GN PEP0655C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId:5833;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE:99021743; PubMed:9804551;
RA Gardner M.J., Reitelin H., Carucci D.J., Cummings L.M., Arayind L.,
RA Koonin E.V., Shalton S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perlman M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RA "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum";
RT Science 282:1126-1132(1998).
RL EMBL: AE001409; AAC71920.1; -
KW Hypothetical protein.
SQ SEQUENCE 558 AA; 66239 MW; 7D1134E0F17A3194 CRC64;

Query Match	8.9%	Score 125;	DB 5;	Length 558;
Best Local Similarity	23.9%	Pred. NO. 30;		
Matches	57;	Conservative	44;	Mismatches 75; Indels 62; Gaps 11.

QY	33	GMHVEKEKNKNDENKRR	---	DEERNKQ	-----	BEHKEIKNHITVTEV	73
		: : : :		: :		: :	
Db	133	GSNSENKDKKKKKKKK	INNNDKKNEISLYLDDG	CFPNDDYDQHLKPIKNE	TEIKN		192
QY	74	KGEAVVKKEAEKLEKVP	--	DVLEMTKALGKITY	--	IVODDIRKHTLSLESDKKI	129
				: :		: :	
Db	193	KSE	-----	QNFPEIQPNNEEKELK	RTFDMNYEELINDNFVEAOV	EEVGEKTKDK	244
				: :		: :	
QY	130	KDIYKDALLEHHVYAA	EGEYEVLYIOSSE	DYVENTERKAL	---	NVYIEIGKILSRDILS	186
				: :		: :	
Db	245	KLIWG	-----	NVQPFYIIPSN	-DYMDAEDMYNDIN	INDININDIND	289
				: :		: :	
QY	187	KINQF	YQKFLDYLNTIK	-NASDSGQDILLTNO	KEHPDPSVEFL	EONSNEVEYEF	242
				: :		: :	
Db	290	KINHITDKITDKINSD	SIDFTSDTDHINKNYKH	-----	NINEDDITF		336
				: :		: :	

RESULT	8	
018244		
ID	018244	PRELIMINARY; PRT; 649 AA.
AC	018244;	
DT	01-JAN-1998	(TREMBLrel. 05, Created)
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)
DT	01-JAN-1999	(TREMBLrel. 09, Last annotation update)
DE	Y57G11C.20	PROTEIN.
GN	Y57G11C.20	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;	
OC	Rhabditidae; Peloderinae; Caenorhabditis.	
OX	NCBI_TaxID=6239;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	McMurray A.;	
RL	Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=94150718; PubMed=7906398;	
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A., Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnson L., Jones M., Keshaw J., Kirsten J., Laister N., Latreille P., Lington J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M., Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R., Smaldon N., Smith A., Sonnhammer E., Staden K., Sulston J., Therry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,	

RA Watson, Welstock L., Wilkinson-Sproat J., Mohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*
RT elegans.";
RT Nature 368:32-38(1994).
PL EMBL: 299281; CAB16521.1; -
DR SQUENCE 649 AA; 76726 MW; 8AE7CDB6D533BA95 CRC64;

Query Match	18.8%	Score 124	DB 5	Length 649
Best Local Similarity	19.6%	Pred. No. 40		
Matches 52	Conservative 56	Mismatches 97	Indels 60	Gaps 8
QY	38	EKEKNKDNKRKDERNKTOEE---HLKEIMKHIVKIE---	KGEVYKKAELKLEK	90
		: : : : : : : : : : : : : : :		
DB	286	EKEHNEOEKKCKNNLEKAEKERHDIHEROAEARAKALBOEKOEKDNEDRERATOK		345
QY	91	VPSPVLMEYKAGIGKIYIVDDITKHISLEALSEDDKKIKIDYGDALHNE-		142
		: : : : : : : : : : : : : : : : :		
DB	346	ERRHLEHRHQAELAKARAKALEKEKKEHRLQOEAKDKDEDEAROKAROKADAREAKSRDYK		405
QY	143	-----YYAKEGYEPVL-----VQGSSEDIVENTEKAL		170
		: : : : : : : : : : : : : : : : :		
DB	406	AEKYAIDPKOVNIPGKYAYASQEAKKILEKYNLEKQERDKIVSKVRESEALIEKAVPRA5		465
QY	171	NVVEYIG-----KILSRDILSKINQYOKFYLVDLNTINASSDGDODLLFTWQLEKHPD		225
		: : : : : : : : : : : : : : : : : : : : : :		
DB	466	SEYHDVALKIPQKPNDRNIYDELKLSFER---ISRIRIVSHPENSNDYL---QFEKQDGR		519
QY	226	FSVEFLFQNSNEVQEVFAKAFAYI		250
		: : : : : : : : : : : : : : :		
DB	520	VSIDFLD--SNILFEFEKSCAFYL		542

Query Match	8.88;	Score 123.5;	DB 1;	Length 880;
Best Local Similarity	23.3%;	Pred. No. 61;		
Matches 70;	Conservative % 50;	Mismatches 94;	Indels 87;	Gaps 15;

ID	ORGANISM	PRELIMINARY;	PRT;	880 AA.
Q9U2C8	01-MAY-2000 (TrEMBLrel. 13, Created)			
Q9U2C8	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
Q9U2C8	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)			
Q9U2C8	PURINE NTPASE, PUTATIVE.			
Q9U2C8	PAB0812.			
Q9U2C8	Pyrococcus abyssi.			
Q9U2C8	Archeae; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.			
Q9U2C8	NCBI_TaxID=29292;			
Q9U2C8	[1]			
Q9U2C8	SEQUENCE FROM N.A.			
Q9U2C8	STRAIN-ORSAY;			
Q9U2C8	Heilig R.;			
Q9U2C8	"Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.";			
Q9U2C8	Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.			
Q9U2C8	EMBL: AJ248286; CAB50131.1; -.			
Q9U2C8	InterPro: IPR003593; AAA.			
Q9U2C8	InterPro: IPR003439; ABC_Transport.			
Q9U2C8	InterPro: IPR001687; ATP_GTP_A.			
Q9U2C8	InterPro: IPR001238; RecF.			
Q9U2C8	InterPro: IPR000533; Tropomyosin.			
Q9U2C8	Pfam: PF00470; RecF.1			
Q9U2C8	PRINTS; PR00194; TROPOMYOSIN.			
Q9U2C8	SMART; SM00382; AAA.1.			
Q9U2C8	Complete proteome.			
Q9U2C8	SEQUENCE 880 AA; 103970 MW; FDB177EC7E026479 CRC64;			

[illegible]

```

07      87 -LLEVPSPVSEMYAIGGIYIVGGDTIKHLSLEALSDKKKIDYGDQALLHEHYV    145
D6      1236 IITEIKKEVKRRVKKRNKNKNKVNIYOELMNEVDYN-----KDTANKGVIEDE--- 1287
07      146 AKEGEPAVLVIOSSEDYVENTERKALN--VYYEIGKILSRDILSKINOPYOR---FLDV 198
D6      1288 -KEREE---VKREVEKEDPVNRKDVGIGOEIIVEEKEEVKKRVKRNKNENKNDVIYGTI 1343
07      199 LNTIKNASDSGQDLLFTNQLKHEHTPDSVFLEONSNEVOEFAKAFYILEPOHR--- 255
D6      1344 MNMEDVNEKDTESDKMKIGREV-----IIEEVEEKVRKRVKNRKNRRRKNE 1393
07      256 --DVL-OLYAPAEFNYMDKEN 273
D6      1394 RKDVIQEIVSEEVENKEKDTKN 1414

RESULT   14
O66878    PRELIMINARY;          PRT: 1156 AA.
066878
A6      066878:
A6      01-AUG-1998 (TREMBLrel. 07, Created)
D6      01-AUG-1998 (TREMBLrel. 07, Last sequence update)
D6      01-JUN-2001 (TREMBLrel. 17, Last annotation update)
D6      CHROMOSOME ASSEMBLY PROTEIN HOMOLOG.
GN      XCP OR AQ.629.
O5      Aquifex acolicus.
O5      Bacteria; Aquificales; Aquificaceae; Aquifex.
O5      NCBI_TaxID=63363;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-VF5;
RX      MEDLINE=98196666; PubMed=9537320;
RA      Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA      Gridman D.E., Overbeek R., Snead M.A., Keller M., Ajay M., Huber R.,
RA      Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RF      "The complete genome of the hyperthermophilic bacterium Aquifex
RF      aeolicus." ;
RL      Nature 392:353-358(1998).
DR      EMBL; AE000659; AAC06839.1; -
DR      InterPro; IPRO03439; ABC_transprott.
DR      InterPro; IPRO01687; ATP_GTP_A.
DR      InterPro; IPRO03405; SMC_C.
DR      InterPro; IPRO03395; SMC_N.
DR      InterPro; IPRO02017; Spectrin.
DR      Pfam; PF02483; SMC_C.1.
DR      Pfam; PF02463; SMC_N.1.
DR      ProSite; PS00211; ABC_TRANSPORTER; UNKNOWN.1.
KW      Complete proteome.
SV      SEQUENCE 1156 AA; 135563 MW; B12DB30F70C0CE49 CRC64;

Query Match      8.6%; Score 120.5; DB 2; Length 1156;
Match Local Similarity 27.5%; Pred. No. 1.2e+02;
Matches 64; Conservative 31; Mismatches 89; Indels 49; Gaps 11.

07      31 DYGMVHVKEKKKKDENKR-KDEERNK-----TOEHKLKEMIHVYKGEAVKKEAA 84
D6      340 EVGTIQLELEKLEKEYKSLEYEREKRLREEBERLKTTFPEVKKLF-----EEK 390
07      85 EKLEKVPVS-----DVLEMYKA-IGGKIYIVDGDTITKHLSLEALSDKKKINDYGDQL 138
D6      391 EKLTERRKSLNKKEQELIORANLNKRIERIKEDINK-----LISPREKIKEIMEKGE 445
07      139 LHETHVYAKEGEPAVLVIOSSEDYVENTERKALN----YELGKILSRDILSK--INOPY 192
D6      446 IKRLKAIRKK-----EEEELRNLTOELINIYEKRLSEVRKKLEELKKGAIEREV 495
07      193 OKFLVIVMTIKNASDSGO-DLLFTNQLKHEHTPDSV-----EPLEQNSNEV 238
D6      496 RSFSVDVDFKDIKGYGSVSELIRVKNPDHITAIEVAGGGRKLFTIVVEDEEV 548

```

RESULT 15

Q26223

ID Q

AC Q

DT 0

DT	0
DE	0

DT 0
DE 0DE
OS
R
P

35
4

OX
E

OX	N
PN	I

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850

RC 5

KC	RX	5	M
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12
13	13	13	13
14	14	14	14
15	15	15	15
16	16	16	16
17	17	17	17
18	18	18	18
19	19	19	19
20	20	20	20
21	21	21	21
22	22	22	22
23	23	23	23
24	24	24	24
25	25	25	25
26	26	26	26
27	27	27	27
28	28	28	28
29	29	29	29
30	30	30	30
31	31	31	31
32	32	32	32
33	33	33	33
34	34	34	34
35	35	35	35
36	36	36	36
37	37	37	37
38	38	38	38
39	39	39	39
40	40	40	40
41	41	41	41
42	42	42	42
43	43	43	43
44	44	44	44
45	45	45	45
46	46	46	46
47	47	47	47
48	48	48	48
49	49	49	49
50	50	50	50
51	51	51	51
52	52	52	52
53	53	53	53
54	54	54	54
55	55	55	55
56	56	56	56
57	57	57	57
58	58	58	58
59	59	59	59
60	60	60	60
61	61	61	61
62	62	62	62
63	63	63	63
64	64	64	64
65	65	65	65
66	66	66	66
67	67	67	67
68	68	68	68
69	69	69	69
70	70	70	70
71	71	71	71
72	72	72	72
73	73	73	73
74	74	74	74
75	75	75	75
76	76	76	76
77	77	77	77
78	78	78	78
79	79	79	79
80	80	80	80
81	81	81	81
82	82	82	82
83	83	83	83
84	84	84	84
85	85	85	85
86	86	86	86
87	87	87	87
88	88	88	88
89	89	89	89
90	90	90	90
91	91	91	91
92	92	92	92
93	93	93	93
94	94	94	94
95	95	95	95
96	96	96	96
97	97	97	97
98	98	98	98
99	99	99	99
100	100	100	100

RA K

RT	h
----	---

RT
YRL
M

DR

50
5

Quer

Best

Matc

?

QY

중

Db

Only

QY

nh

18

05

10

Qy	216	TNQLKEHPTDSEVPELEQNSNE	237
		:::	:: :: :
Db	382	TDKIKAEVSDKELKKCEQSFND	403

Job time: 513 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:50:38 ; Search time 72.22 Seconds
(without alignments)
820.601 Million cell updates/sec

Title: US-09-747-521-2_COPY_1_778

Perfect score: 3987

Sequence: 1 MNKKKEFIKIVISMCLVTAL.....TSYGRTEAEFFAEAFRLMH 778

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3987	100.0	809	1 J00032	anthrax toxin leth
2	501	12.6	800	1 B58106	hypothetical prote
3	497	12.5	800	1 J50029	adenylate cyclase
4	255	6.4	2269	2 T28677	thoptry protein -
5	253	6.3	1979	2 C71622	hypothetical prote
6	253	6.3	2401	2 T28676	rhophtry protein -
7	239	6.0	1127	2 T28317	ORF MSY156 hypote
8	235	5.9	3119	2 T18414	protein 9377 - mal
9	233.5	5.9	1939	2 T18372	repeat organelle
10	232	5.8	2829	2 A42771	reticulocyte-bind
11	229.5	5.8	946	2 S28061	SCP1 protein - rat
12	228.5	5.7	993	2 S49461	synaptonemal compl
13	224.5	5.6	1558	2 T18403	RESA-H3 antigen p
14	222.5	5.6	3724	2 T18427	hypothetical prote
15	220	5.5	2166	2 G70163	hypothetical prote
16	219.5	5.5	2663	1 S28261	centromere protei
17	218.5	5.5	1163	2 D64315	type I restriction
18	218	5.5	980	2 E71606	hypothetical prote
19	217	5.4	1088	2 T18559	hypothetical prote
20	215	5.4	2245	2 T18278	myosin heavy chain
21	214.5	5.4	1313	2 F36673	hypothetical prote
22	213	5.3	1516	2 E71619	RAD2 endonuclease
23	209.5	5.3	1005	2 A64465	hypothetical prote
24	209.5	5.3	1128	2 G66266	hypothetical prote
25	209.5	5.3	1780	2 T17272	hypothetical prote
26	209.5	5.3	1957	2 T38077	hypothetical colle
27	209	5.2	1679	2 S48385	hypothetical prote
28	208.5	5.2	691	2 S48390	hypothetical prote
29	208	5.2	1230	2 S56850	SMC1 protein homol

30	206.5	5.2	1650	2 T18444	hypothetical prote
31	205.5	5.2	756	2 C64236	protein V (fcv) h
32	205.5	5.2	1875	2 S38173	myosin-like protei
33	203.5	5.1	2139	2 T18296	myosin heavy chain
34	203	5.1	3225	2 I52300	giantin - human
35	203	5.1	3259	1 A56539	giantin - human
36	202	5.1	2710	2 A37052	toxlin A - Clostrid
37	201.5	5.1	841	2 A86188	hypothetical prote
38	201.5	5.1	1909	2 A45592	liver stage antigen
39	201.5	5.1	1922	2 T00637	hypothetical prote
40	201	5.0	1790	2 S67593	transport protein
41	200	5.0	845	2 I48176	synaptonemal compl
42	200	5.0	2485	1 H71621	serine/threonine-s
43	199.5	5.0	1252	2 B42771	reticulocyte-bind
44	199.5	5.0	2116	2 A26655	myosin heavy chain
45	199	5.0	1312	1 BMBIDL	RAD50 protein - ye

ALIGNMENTS

RESULT 1
J00032
anthrax toxin lethal factor pXOI-107 precursor - Bacillus anthracis virulence plasmid
C:Species: Bacillus anthracis
C:Date: 31-Mar-1990 #sequence_revision 11-Nov-1994 #text_change 11-May-2000
C:Accession: J00032; C59104
R:Bragg, T.S.; Robertson, D.L.
Gene 81, 45-54, 1989
A:Title: Nucleotide sequence and analysis of the lethal factor gene (lef) from Bacill
A:Reference number: J00032; M01D:90034185
A:Accession: J00032
A:Molecule type: DNA
A:Residues: 1-809 <BRA>
A:Cross-references: GB:M29081; NID:9143143; PIDN:AAA79226.1; PID:9143144
R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koeh
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pXOI, the large Bacillus anthracis plasmid harb
A:Reference number: A59091; M01D:99445483
A:Accession: C59104
A:Molecule type: DNA
A:Residues: 1-809 <OKI>
A:Cross-references: GB:AF065404; NID:94894216; PIDN:AAD32411.1; PID:94894323
A:Experimental source: strain Sterne
A:Note: similar to Anthrax toxin lethal factor precursor; lef, plasmid pXOI, B. anthr
C:Comment: This lethal factor of Bacillus anthracis is part of the tripartite protein
her they cause anthrax, an infectious and often fatal disease of cattle, sheep, and o
C:Genetics:
A:Gene: lef; pXOI-107
A:Genome: plasmid
C:Superfamily: anthrax toxin lethal factor; lethal factor amino-terminal homology
C:Keywords: toxin
F:1-33/Domain: signal sequence #status predicted <SIG>
F:34-809/Product: anthrax toxin lethal factor #status predicted <MAT>
F:44-295/Domain: lethal factor amino-terminal homology <LFA>

Query Match 100.0%; Score 3987; DB 1; Length 809;
Best local Similarity 100.0%; Pred. No. 1.3e-152;
Matches 778; Conservative 0; Mismatches 0; Indels 0;

QY	1	MNKKKEFIKIVISMCLVTATITSGPFIPLVQAGAGGDMVKKKKKKDKKKRDEE	60
DB	1	MNKKKEFIKIVISMCLVTATITSGPFIPLVQAGAGGDMVKKKKKKDKKKRDEE	60
QY	61	RNKTQEHLEIKKHIVKIEVKGEEAVKKEAEKLEKPSDYLENYKAIGKTIYVDGD	120
DB	61	RNKTQEHLEIKKHIVKIEVKGEEAVKKEAEKLEKPSDYLENYKAIGKTIYVDGD	120
QY	121	IRKHISLEALSDKKIKIDYKDALLHEHYVYAKGEPVLVYOSSEDEVTEKALNV	180
DB	121	IRKHISLEALSDKKIKIDYKDALLHEHYVYAKGEPVLVYOSSEDEVTEKALNV	180

QY 181 YVEIGKILSRILSKINOPYOKFLDVLTINKNASDSGQDILLFTNOLEKHPDSEVFLE 240
DB 181 YVEIGKILSRILSKINOPYOKFLDVLTINKNASDSGQDILLFTNOLEKHPDSEVFLE 240
QY 241 QNSNEVOEFAKAFAYIEPOHRDVLQYLAPEAFNMDKFNQEOEINLSLEELKQRMLSR 300
DB 241 QNSNEVOEFAKAFAYIEPOHRDVLQYLAPEAFNMDKFNQEOEINLSLEELKQRMLSR 300
QY 301 YEKMEKIKOHQWMSDSISEEGRLKKLQPIPEPKKDDIHSLSQEEKLLKRIQIDSS 360
DB 301 YEKMEKIKOHQWMSDSISEEGRLKKLQPIPEPKKDDIHSLSQEEKLLKRIQIDSS 360
QY 361 DFLSTEEKEFLKQIQRDLSSEEEKELLNRIOVDSSNPLSEKEKEFLKAKIDQPYD 420
DB 361 DFLSTEEKEFLKQIQRDLSSEEEKELLNRIOVDSSNPLSEKEKEFLKAKIDQPYD 420
QY 421 INQRLQDTGGLIDSPSLNLDVRKQYKRDIONIDALLHOSISTLYNKYYLENNINNL 480
DB 421 INQRLQDTGGLIDSPSLNLDVRKQYKRDIONIDALLHOSISTLYNKYYLENNINNL 480
QY 481 ATLQADLVSTDNTRKINRGIFNEFKKNEKYSISSNMYIVDINERPALDNERLKRIQLSP 540
DB 481 ATLQADLVSTDNTRKINRGIFNEFKKNEKYSISSNMYIVDINERPALDNERLKRIQLSP 540
QY 541 DFRAGYLENGKLLIORNGLIKRDOYIIKOSEKEXIRDAVVRKSKIDTYIOEAOLIN 600
DB 541 DFRAGYLENGKLLIORNGLIKRDOYIIKOSEKEXIRDAVVRKSKIDTYIOEAOLIN 600
QY 601 QEMKALGLPKYTKLITFNHNRASNIVESAYLILNEMKNNISDLIKKATNYLVQNG 660
DB 601 QEMKALGLPKYTKLITFNHNRASNIVESAYLILNEMKNNISDLIKKATNYLVQNG 660
QY 661 RFVETDTLPIAIOYRHODEIYEQVHSKGLYVESRILHGPSKGYELNDSBGFHE 720
DB 661 RFVETDTLPIAIOYRHODEIYEQVHSKGLYVESRILHGPSKGYELNDSBGFHE 720
QY 721 FGAHVADYAGLLDKNOSDLVTNSKKFFIDIFKEGSNLTSGRTNEAEFFAEFLRH 778
DB 721 FGAHVADYAGLLDKNOSDLVTNSKKFFIDIFKEGSNLTSGRTNEAEFFAEFLRH 778
RESULT 2
B59106
hypothetical protein px01-122 - Bacillus anthracis virulence plasmid px01
C:Species: Bacillus anthracis
C:Date: 12-Nov-1999 #sequence, revision 12-Nov-1999 #text_change 09-Jun-2000
C:Accession: B59106
R:Oklinaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Kelm, P.; Koehler
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of px01, the large Bacillus anthracis plasmid harbored
A:Reference number: A59091; MUID:99445483
A:Accession: B59106
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-800 <OKT>
A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AMD32426.1; PID:g4894338
A:Experimental source: strain Sterne
A:Note: Similar to calmodulin sensitive adenylate cyclase, edema factor, cya, plasmid pX
C:Gene: px01-122
A:Genome: Plasmid
C:Superfamily: calmodulin-sensitive adenylate cyclase; calmodulin-sensitive adenylate cy
P:34-286/Domain: lethal factor amino-terminal homology <LFA>

Query Match 12.6%; Score 501; DB 2; Length 800;
Best Local Similarity 24.6%; Pred. No. 3.6e-13;
Matches 202; Conservative 145; Mismatches 281; Indels 194; Gaps 36;

QY 4 KKEFI----KYISMSCLVATITLSPGVFIPLVQAGGHDVGMVKEKKNKDKENKRD 59
DB 3 RNKFIPIKMFSTISPSVLLFAISSQALEVNA-----MNEHYTSSDIKRNHKT 49

QY 60 ERNKTQEHLEKIMKHIVKIGVEAVKKAERKLEKPSVDLEMKATIGKITIYVDC 119
DB 50 EKNTEKKEFKDSINNLYVKTETFTEDLKIQOTDOLLKIPKVLSEYSGEYTFDD 109
QY 120 DITHISLEAISEDKKKIKDIDYGDALEHHYVAKEGEYEVLTIOSEEDVENTERALN 179
DB 110 DLVEHKELODLSSEEEKSNMNGEKEVPASRFVEKKEKREPKLII-NIKDVAINSSEOSKE 168
QY 180 YVEIGKILSRILSKINOPYOKFLD--VLNTINKAS--DSDGQDILLFTNOLEK---HPT 232
DB 169 YVEIGKILSRILSKINOPYOKFLD--VLNTINKAS--DSDGQDILLFTNOLEK---HPT 232
QY 233 DFLSTEEKEFLKQIQRDLSSEEEKELLNRIOVDSSNPLSEKEKEFLKAKIDQPYD 292
DB 224 SIDINFTEKENLTEROFHAFSLAFSYFFAPDHTVLELYAPDMFEYENKLU----- 271
QY 293 KQWMLSRYEK--WEKIKOHQWMSDSISEEGRLKKLQPIPEPKKDDIHSLSQEEK 350
DB 272 -----EKGEFKI-----SESLKKG-----YKDRIDVL-----KGEKA 301
QY 351 LKRIQIDSSDLSSTEEKEFLKQIQRDLSSEEEKELLNRIOVDSSNPLSEKEKEFLK 410
DB 302 L-----KAGLVPEHADARKI-----ARLNTYILFRVYNKLTATN---LI 339
QY 411 KLRKIDQPYDINQRLQDTGGLIDSPSLNLDVRKQYKRDIONIDALLHOSISTLYNK 461
DB 340 KSGVATKGLNVHGKSSDGPVAGYIPDQDLSKRGQOLAVENKLENLEKKSITHEGEIG 399
QY 462 STLYNKIYL-YENNNINNLATLQADLVSTDNTRKINRGIFNEFKKNEK-----YSISSN 515
DB 400 -----KIPKLDHLRIEELK-----ENGILKCKKEIDNCKKYYLLESN 438
QY 516 YMIVDINERPALDNERLKRIQIOLSPDTRAGYLE-NGKLIILORNGLEIKDOY-IIKOSEK 573
DB 439 NQVVEF--RISDENNEVOYK-----TKEGKITVLGEFFNNRNIEVMAKNVEGYLPLTA 490
QY 574 EYIPIADKVVYKSIDRPIQEOALINQENMKAVGLPYTK-----LITFNHNRAYA 625
DB 491 DY---DLFALPSSLTEIKKQIPQ-----KWDKVVNTPSLEKQKGVNLTLYKYEIERK-- 541
QY 626 SNIVESAYLILNEMKNNISQDLIKKATNYLVQNGRFEVDTLPIAIOYTHOD-EIY 683
DB 542 ---PDSTKGLTSMNQKQMLRDL-----NEAVYTYGTGADVNVHTEGDNEBFP 587
QY 684 EQVHSKGLYVESRILHGPSKGYELNDSBGFHEGHAVDYA-----GYL----- 732
DB 588 EKDNIEFLINEGEFIIILKNMEMTGRFTEKNTGKDYLYFNRSYKNKIAPGNKAYIEWTD 647
QY 733 -LDKNOSDLVTNSKKFFIDIFKEGSNLTSGRTNEAEFFAE 773
DB 648 PITKAKIWTIPTSAEFL-----KNLSSIRSSNVGVYKDS 682
RESULT 3
J50029
adenylate cyclase (EC 4.6.1.1) precursor, calmodulin-sensitive - Bacillus anthracis
A:Alternate names: anthrax toxin edema factor
C:Species: Bacillus anthracis
C:Date: 31-Mar-1992 #sequence, revision 11-Nov-1994 #text_change 19-Jan-2001
C:Accession: J50029; PS0307; J50602
R:Robertson, D.L.; Tippetts, M.T.; Leppla, S.H.
A:Title: Nucleotide sequence of the Bacillus anthracis edema factor gene (cya): a cal
A:Reference number: J50029; MUID:89211974
A:Accession: J50029
A:Molecule type: DNA
A:Residues: 1-800 <ROB>
A:Cross-references: GB:M24074; NID:g142812; PIDN:AAA79215.1; PID:g142813
A:Accession: PS0307
A:Molecule type: protein
A:Residues: 34-48 <RO2>
R:Escuyer, V.; Duflot, E.; Sezer, O.; Danchin, A.; Mock, M.
Gene 71, 293-298, 1988

QY 350 -----ELKRIQSDSDFLSTE-----EKPEFLKQIDIRSDISEEKFL-----LN 391

Db 1650 CMLFHTTTLAEFLKIKIDYKSFIESATFKSEFLKYG-DTSSNLNDIATLQKLYDH 1708

QY 392 RIQ-----VDSSNPLSEKEFEFLKRLKLDIQPYDINQRLDTGGGLDPSINDV 441

Db 1709 QINKYVTSKLSLDATNNDNLIEKE-----ATQAIKRLKFLTIDSNMIDA 1755

QY 442 RQYKRDIONI--DALHHOSTGS--TLYKKIYLENMNI----- 476

Db 1756 MALHNKIQMVFENSELHKSIESIKQLYKKMHVFKLNLGQINGKYPDISKQFNIIQLQ 1815

QY 477 -NNLTATLGLADY--VDSDNTRKINRGI-----FNEFKNKFKYSISSNYMIVDNE 523

Db 1816 ESELTRANLN-DLKEIGQKISDKKNNFLHALNETPIRPFNTLKEIYHDYKRYQIDEIEN 1874

QY 524 RPALDNE-----RLKWRIQLSPTTRAGYLENGKLIILQRIIGLEIKDYOIIKQSEK 573

Db 1875 ITSEENNETTLYIDTFTTKLEKQVOSILNFVTTEYDENSNIKQIHODITNENDVSKIKESL 1934

QY 574 EYIRIDAKVVR-----SKIDRKIQQAQNLINQEW--NKAGL 609

Db 1935 TTIIQSFOELNKINGIKAKQFYDNNNNINNTPTISIDQVDVKKHISKDLTTEHELLEI 1994

QY 610 PYRTKLI--TFNV--HNRYASNIVESAYLLINEMKNNIQ--SDLIKVTNYLVQGN 659

Db 1995 QKSLIEDIKNSTYEINGANNNYNITRNYEQOTNKIQNNNSMDELDIILQKLYNKE-- 2052

QY 660 GRFVFDTIDLPINAEQYTHODEIYEQVHSGKLYPESSRSLILHPSKV-----ELRND 713

Db 2053 -----SETKLPITGNKNKNTVTSIISRI-NKVYNLIESE---YGNNNVSYVAKKEED 2102

QY 714 SEGFTHFGHAVDVAGYLLDKNOS--DLVYNSKKFDIFK-----EEGSNLTSGERT 764

Db 2103 ANSIILD-----LDKSQNLKLDLIQNLKIIDLDKKNKQETIENRNLIQTINRE 2150

QY 765 NE 766

Db 2151 QE 2152

RESULT 5

C71622

hypothetical protein PF00145c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000

C:Accession: C71622

R:Gardner, M.J., Tettehin, H., Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; Perlman, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O. Science 282, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A11600; MUID:99021743

A:Accession: C71622

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1979 <GAR>

A:Cross-references: GB:AE001375; GB:AE001362; NID:g3845105; PIDN:AACT71819.1; PID:g3845105

C:Experimental source: clone 3D7

C:Genetics:

A:Gene: PF00145c

[illegible]

A:Cross-references: GB:M34281

Query Match 6.3%; Score 233; DB 2; Length 2401;

Best Local Similarity 20.2%; Pred. No. 0.012;

Matches 182; Conservative 159; Mismatches 340; Indels 222; Gaps 38;

```
QY 2 NIKKEPIKYSCLVTAITLSPVPIPLVVGAGHGVDGMHKEKKKDEKRRDEER 61
  |||||||
DB 1389 NIYEPIKSYDLITHLEVSKEPIYEDQIKKRITAQNELTNINAKASYLDDIER 1448
QY 62 NKTQF--EHLKEIMKHI-----VKIEVGEE-----AVKKEPAKLEKVPDVL 104
  ||:|||||
DB 1449 NEFDRIYTHFKKLNVDNKFNTNEYSKVN-KGFQDNISNSINNVKSTDEMLNLINQKR 1507
QY 105 EMYKAIGKIYI-----VDG-DITRHS---LEALSEDK 134
  |||||
DB 1508 EMYAMIVSKRYYSYKAEENIFINIPKLANSLNIQKSSGIDLFKNINAIPLYLDSQK 1567
QY 135 KIKIDYGDALDLEHYVYAKGEYEPVL-VIOSSEDEVNTEKALNVYEIGKILSRDL 193
  ||:|||||
DB 1568 KOTLPIFPEPEKTSERYTISDSTYNLIDILKRSOELQKKEOALNLIFE-----NRLH 1622
QY 194 SKINOPYOKFLDVLNTIKNASDSGDGLFTNQLKEHPTDFSVFELEONSNEVOEFAPKA 253
  ||:|||||
DB 1623 DKV-QATNELKDTLSDLKRRKEQ-----ILNKVK-----LLHKSNEMLKLSGNS 1666
QY 254 FAY--YIEQHDVLDQLVPEAFNFMKNEOEINLSLELDQRLSREKREKIKOHY 311
  ||:|||||
DB 1667 QWYDTLESSEKYDKIK--EKSNNYEKEKENLGINDFVAMEBO-FNNDIKDEKLENNY 1722
QY 312 OHMSD---SLSEGRGLL---KKLO-----PIEPK---KDDIHSLOEKE-- 350
  ||:|||||
DB 1723 KISEKDNVFESENNNIILOSKKKLKELTANFAEIKKIEDKIIEKGLINKLETTRKDCM 1782
QY 351 -----LAKRIQIDSSDF-----LSTEEKEFLKLQIDIRSLSEEEKEL-----LNRI 393
  ||:|||||
DB 1783 LFTYKLVETLKIKTDTYKFTSATKFSKEFLKYIDA-TSNSLNDINDMTLQKRYDLNOI 1841
QY 394 Q-----VDSNPLSEKEKEFLKRLKLDIQPYDINORLQDPTGGLDPSINLDYKR 443
  ||:|||||
DB 1842 NRVASWADATNDNNNLEKEKEATKTINNTLEF-----TDSNKLIDADGLH 1890
QY 444 QYKRIQIONIDALLHOSIGS--TLVYKITYLENNMI-----NNTLATGADL 487
  ||:|||||
DB 1891 NKKIQIIVNSLHKSIDSISKQLYKMAFKLLINIGHINKKYPDISKEPDNILOESEL 1950
QY 488 VDSNTKTINRGIFNEKKNFYSSISSNMYIVDINERPALDNERLKMRIQLSPDRAGYL 547
  ||:|||||
DB 1951 TANLNDLKEIGOKISDKKQFLHALSET-PIPNFNTLKEIYHDIVKYKNOIDEIENITNE 2009
QY 546 ENGKLLIONIGLEI-----KVOIHKOSEKEIYRIDAVVPKSKIDTKI 592
  ||:|||||
DB 2010 ESENTTLYMDITLTKLKKVQOSILNEVTYENDSNIIOHIONNENDVSKI-KESLETTI 2068
QY 593 QEAQLININQEMKALGLPYTKLITFNHRYASINIVESAYLLINEMKNNOISQDLIKKYT 652
  ||:|||||
DB 2069 QSFQKRLN-----KLNEIKAFYDNNNNINNVISITISQDVNDVKKHISDL--TIE 2116
QY 653 NTLVDNGRFVFTDITLPLNIAEOY--THODEIYEOVHSGLYVPESRSILLHGSKGVEL 710
  ||:|||||
DB 2117 NELLIQ-----IQLSLEDIKKSTYDIRSEQITKYVNPJHDVEQOTKRIQNPNK---- 2165
QY 711 RNDSEGFIEHFGHADVAGYLLDKN-QSDL---VTNSK-----KRIDFKKE 754
  ||:|||||
DB 2166 -----DEIDDLIQEIVNKNKESLKLPTIINKNDVNTPIISRIDKVINLIKSE 2213
QY 755 GSN 757
  ||
DB 2214 YNN 2216
  ||
```

RESULT 7

T28317

ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus

C:Species: Melanoplus sanguinipes entomopoxvirus

C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000

C:Accession: T28317

R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.

J. Virol. 73, 533-552, 1999

A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.

A:Reference number: 220484; MID:99102612

A:Accession: T28317

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1127 <AFO>

A:Cross-references: EMBL:AF063866; NID:g4049647; PID:AMC97677.1; PID:g4049717

A:Genetics:

A:Note: MSV156

Query Match 6.0%; Score 239; DB 2; Length 1127;

Best Local Similarity 20.8%; Pred. No. 0.017;

Matches 159; Conservative 121; Mismatches 252; Indels 234; Gaps 35;

```
QY 60 ERNKTQEE--HLKEIMKHIYKIEVGEEAVKKEAEKLEKVPDVLKYKAIGKIY- 115
  ||:|||||
DB 24 ENKVSIDITNSLYEILNNT-----KFSDKITNELIKYKNIYKIEYIF 66
QY 116 -----IYGDITKHSLEALSEDKKKID-----IYGDALLHEHYVYAKGEYEP 161
  ||:|||||
DB 67 MHNQFINVNIILQYLVIEENNEIKNCIKENKPPCKNPLY-----NITYRKKKIYID 118
QY 162 LVIOSSEDEVNTEKALNVYE-----IGKILSDISKINOPY-----QKFLDVL 207
  ||:|||||
DB 119 L-----DYEKKDKELVINIEQKNAVDTK--NDIKNNVNIHSDNETIITGETLIDIL 170
QY 208 NTKNASDSGDGLFTNQLKEHPTDFSVFELEONSNEVOEFAPKAFAYIIEPQHDVLD 267
  ||:|||||
DB 171 NKLKLVSSDEKOLI--EQIYKINNKKEIF--KNIDVQKEIKF----- 211
QY 268 LVPEAFNFMKNEOEINLSLELDQRLSREKREKIKOHYOHMSDLSSEEGGLK 327
  ||:|||||
DB 212 -----KQELNKLNLDL-----SKREFK----- 229
QY 328 KLOPIEPKKDDIHSLOEKEELKRIQIDSSDFLSTEEKEFLKLQIDIRSLSEEEK 387
  ||:|||||
DB 230 -----KQELNKTIDKQOELIKKNDKEINFPINDEKQLDQINSKI-NFLWENIK 280
QY 388 ELINRIQVDSNPLSEKEKEFLKRLKLDIQPYDINORLQDPTGGLIDPSINLDYKRQYKR 447
  ||:|||||
DB 281 GYVNN-LYTEKKNKISNLONEILNK-DSTIKSLDEKQKILD-----ELDK 322
QY 448 DIONIDALLHOSIGTLVYKITYLENNMNTNLTATGADLVDSNTKTINRGIFNEKKN 507
  ||:|||||
DB 323 NINNT-----TSLYKN-----SNKTTNIQOQLLESSLTD-FNNANIN---INELSK 365
QY 508 FKYSISSNMYIVDINERPALDNERLKMRIQLSPDRAGYLENGKLLLORNIIGLEIRVOI 567
  ||:|||||
DB 366 IK-----LFD-NIOQLKNNNDITEQNNKIT-----DFPNNSTRIFKEKLDIYEVKID 411
QY 568 IKOSEKEYIRIDAKVVPKSKIDTKIQEAQLININQEMKALGLPYTKLITFNHNR---- 623
  ||:|||||
DB 412 IKNNNLOKLEESYK-----KIDQOTEYKKNKINKEVNDIIEI-LKNNLOKLEENKIDE 465
QY 624 ---YASINVESAYLLINEMK-NNIQ-----SLLIKVTYVYLDGNRQFVFTDITL 670
  ||:|||||
DB 466 QTEYKKNKINKEVNDIIEIKNNNLOKLEENKKNINOKLTKLNDISNTELF-----NKL 520
QY 671 NIAEQYTHODEI-----YEOVHSGKGLYVPESRSILLHGSKGVELRNDSEGFIEHFGHA 724
  ||:|||||
DB 521 NISDPKDKSREIAKMLNTEYQL-----RKDLLENINKTNELMKLSDNKLSLEOL 570
QY 725 VDDYAGYL-LDKNQSGLVYNSKKFIDIFKEEGSNLTYSRTNEAE 768
  ||:|||||
DB 571 YDSKKNILDGIDIKYNSLAKKNDKIDYF-----SNIEKFDIYVIE 612
```

RESULT

8

Tl8414
protein g377 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
Accession: Tl8414
R:Handman, E.; Osborn, A.H.; Symons, F.; Van Driel, R.; Capral, R.
Mol. Biochem. Parasitol. 74, 143-156, 1995
A>Title: The Leishmania promastigote surface antigen 2 complex is differentially expressed during infection
X:Reference number: Z18933; MUID:96360472
A:Accession: Tl8414
A>Status: Preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-3119 <NAN>
A:Cross-references: EMBL:L04161; NID:g309687; PID:g309688; PIND:AACJ7257.1

Query Match 5.9%, Score 235; DB 2; Length 3119;
Best Local Similarity 19.8%; Pred. No. 0.088;
Matches 174; Conservative 149; Mismatches 301; Indels 254; Gaps 43;

OY 27 FLPIVQSGAGGSD-----VGCMHYKEKKNNDENKRKDDEENKQOEHLKIKMHIVKIE 80
DB 964 FTISEGEHNDRRLRKRIEANKLEWKKRRNEOOREKKAAEE--DMMETIQKH 1021
OY 81 VKGEAVAKKAELKLEYSPDLVMYAIGKITIYDG----DTTKHISLSALSED--K 134
DB 1022 METSKLEEVEDEVTOODEFD-----KYAIIDOBELEFLTRDSEGSBDVPK 1071
OY 135 KKIK-----DIYGKALLHEHVVAKEGYEPVLIOGSEDDY--ENTEKALNVYYEI- 184
DB 1072 DKVKFPFGGRSPPDSYYYNTAISFFPEKKMEELYN--TSISSLVNYKETNRPFDDVIYEKL 1129
OY 185 -----GKILSRDILSKIQQ-----PYQ-----KFLDVLTINKNASPSGD 218
DB 1130 SKTYPKFDDLTSQTGTCKNCNKLFQRKLETITDKEXOKNIQSRYKVVIDILDIIQ--KKANG 1167
OY 219 QDLFTNLQKEHPDFSVEPLEONSNEVOEVFAAF----- 254
DB 1188 KYIIIONLIETIDIYUKDVVRSLDRFKFYKNFRKYLGRKKMKMLDFRAOFKGARFIKD 1247
OY 255 --AYTIFEOHRVDLO-LVARPAFNMDKENFOEINLSLEEKDOBMISR---YEWEEKIK 308
DB 1248 LTTYMEBETYKTVELDIYMKKKKYKEES-----KMRTIISNDIYEYNKOIK 1296
OY 309 OHYQMWSDSLSEEGRGLLKQLQIPREPCKDDIHSLSQEKKELLKRIODSSDELSTEEK 368
DB 1297 EHY-HKVPTISEH-----KFEQIRQHMRDKIENTIELLYEKMYVQOIDLTINY----- 1344
OY 369 FELKLQIIDIRDSLEEKELLNRIQOVSSNP-----LSEKEPEPLKAKLDOIOPYDN-- 422
DB 1345 -----HQLENHSSELLQAOLQONKNIPRHILNVLEKKILETTKRKKRN-KP-DISTS 1391
OY 423 ----ORLODTGCGLDSPINLDVRKQYKRDIQNIDAIALHQSIGSTLYNKIUYLEANN- 475
DB 1392 SHATDEGVSDPR-LIBRAHHNGDIGKBEND-----EVLILLQIOSL--KTMGMGNQ 1442
OY 476 ----INNLTATLGADLVDS-----TDNTKINGICINFEEKNFYSISTSNMTIVDNER 524
DB 1443 VGSILEKLNINS-----DOYOLLQDRKLNVVEDIYKTL-RNFEXHYTELHKESKIRE 1493
OY 525 PALDERLERLKWRITQSPTRAGYLENGKLLIORNISGLEIKDYQ-----TIKOS 571
DB 1494 KFI-----TKDVDLSNVYSTLEIMYAKFLLDHOEMSFEDELEKILYLEEE 1539
OY 572 KEKYRIDAKVVPRKSIDTKIOEAQ--LNTNQEWNAKLAGLPKYTKLTFVNUNRVASNI 628
DB 1540 RKKYITTEIQI--RDLTSLTIONGGDHNNNNNN-----NVRNNLKKQV 1583
OY 629 VESAULLINEKKNNI-QSDLIKKVT-----NYLVGNGRFVFPTDITLRPIAADQYTHODEITY 683
:: : ::
:: : :: :

```

D0 1584 LKDLLEISLKGHILEVIRKKNIALEQINVTNN-----TNEIVDVIADLMAPARIY 1637
Q0 664 ---EQVHSGGLVYPPESRSLILHGSPKGVGLRNDSEGFTHFPGVAVDYAGYLIDKQSD 740
D0 1638 PVSEDIYDITVIVRDTAA-----INNTLRHFVTFDQIYDIDHL-----F 1681
Q0 741 VTNSKRFIDIFEK---EGSNLTSYGRTNEAEFFAAEF 774
D0 1682 VYNIKEL--TYKENLADEKYNANYEENRFLHLLEEF 1717

RESULT
9
118372
repeat organelle protein - Plasmodium chabaudi
GSpecies: Plasmodium chabaudi
GDate: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
CAccession: T18372
RWarner, E.B.; Taylor, W.R.; Holder, A.A.
Mol. Biochem. Parasitol. 94, 185-196, 1998
ATitle: A Plasmodium chabaudi protein contains a repetitive region with a predicted
AReference number: Z18922; MUID:98418765
AAccession: T18372
AStatus: preliminary; translated from GB/EMBL/DBJ
AMolecule type: DNA
AResidues: 1-1939 <WR>
ACross-references: EMBL:U03145; NID:g1151157; PID:g1151158; PIDN:A063403.1

Query Match 5.9%; Score 233.5; D2: Length 1939;
Best local similarity 19.4%; Pred. No. 0.05;
Matches 187; Conservative 170; Mismatches 280; Indels 325; Gaps 48;

Q0 36 GHGDV--GMHVKEKKNKDNKRKDEERNKQOEHLKEIM---KH---YKIEVKG-- 83
D0 898 GHREVVAGLEEKKKEVAVLEEKHKKEIAKLEGH--KEYVAELGEKKKEVAVGLEAKHNL 956
Q0 84 EFAVYKKAEEKLKEVPSVL-----EMYKAIGKIYIVGDITK-----HISLEALSDDK 135
D0 957 EEGHKEMVAE--LEKRNADLVAVLEEQHKA-----EITKLEEKKEVAVAGIEEKY 1004
Q0 136 KIKDIYGDALLHEHYVAAKEGEVPLVYQSSSEDEVVENTEKALNLYYEIGKILSRDIL 195
D0 1005 KVEAL---KLAEEH---KD-----VVTKLGQHKKEIAKLEDDGHKEVYNEVEKKNASL 1051
Q0 136 INOPAKPELDVNTIKNASDSGQDILFTNOLKENHTDSVEPLEDONSNEVDGVFPAKFA 255
D0 1052 LMLLEENHNKNEIKLKEHKESASDLVEKLYOKDEEVKSNMKIEELTVYIKDL----- 1105
Q0 256 YVIEQHRDVLQVAPAFNYMDKFE--OEIN---LSLEELKD----- 294
D0 1106 -----NDSIMCYKQILIEEYKREYNEEINKLIYONEMKDMNDKILKEKENIKKL 1158
Q0 295 QRMLSRYEKWEKIKOHYOH-----WSDSLSE--EGRGLLKKLIQIPIE 334
D0 1159 NKKLSNYKVFETKEWTKSEMVVENKERRIIVDSCKENISSESDVEGKGGLMKTLISL 1218
Q0 335 PKKDDIHSLSGEKELLKRIQDSDFLSTEEKELKIQI----- 376
D0 1219 -KKEKNITSINDKNE-----SELVDITSAIINKIEEMKKEIKEDNGKNIEDLKN 1269
Q0 377 -----DIRDSLEEEKELNRIQVDSNPLSERK-----EFLK-----KLKL 414
D0 1270 ILDLSELNILENNMKNVLTDENNNLKKEIEI--KDNKLNEKKNEMTEILINDIILK 1328
Q0 415 DIQPDINDRLDITGGLDPSINL-----DYAKQYKRIQNIQDALLHOSIG--STLYNK 467
D0 1329 EISEW-----KDEEKLTKENIKLNDIEQINKEKRIEEMLMIKFENINEINVTSLKQ 1382
Q0 468 IYL-----YE-----NMNI--NNLTATGLADIVDSTON--TKIRGIFNE- 503
D0 1383 IEIEKMAKLEELKNTELLAEKRETMGTSINDNDKIVENKNIIEEDTSKQNNLKNKVEDKT 1442
Q0 504 ----FKKNFYKSISSNYMIVDINERPALDNERLAKWRIQLSPDTRAGYLENGKLILQIRNI 558

```

```
Db 1443 GDDINCEKNNDAQKEISYLKDEIKKISMLYGEEL-----NRKNKYDEKVK-----NL 1489
QY 559 GLEIKDVOI-----IKOSEK-----EYTRID 579
Db 1490 TNELKELIKNNKKEEELIAMELNKLNKIKENKNSVKONDESSNNITTKDDKPEYVNSD 1549
QY 580 AKVVP--KSKIDTKIOEAQNLINOENMKALGLPKYTKLTFNVHNRAVSIVESAIVILN 637
Db 1550 DKIQOKMAVLVYLKLEKP-----DLMDNINSLEK-----ENFRVMSIVVE----- 1590
QY 638 EMKNNTIOSDLIKKVTNYLVLDNGNR-----FVFETDI--TLPNIAEQYTHODEIYQVH 687
Db 1591 --NKVNQNDKIYGIYSYFKKCEKELKNDMLYICLVYKLDLISLEFLNDNFYNLEPKIDKIL 1648
QY 688 SKGLYVP--ESRSTILHGPSKGVLELRN---DSGCFI---HEFGHAV-----D 726
Db 1649 WKQMYIPTETIRILFLRFYSFLDKLRNVKCVNEEYNNERYEYSWALFQTYLETASMLKK 1708
QY 727 DYAGYLLDKNQSDLYVNSK-----KPIDI-----FKEGSNLT 759
Db 1709 EMYYVLEKAEKDCSCENSSNFPDKPTITDILNFSKDSIRLKTIAQLKELNFEREAKNLT 1768
QY 760 SY 761
Db 1769 NY 1770

RESULT 10
A42771
reticulocyte-binding protein 1 - Plasmodium vivax
C:Species: Plasmodium vivax
C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 28-Apr-1995
C:Accession: A42771
R:Gallinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
Cell 69, 1213-1226, 1992
A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
A:Reference number: A42771; MUID:92315338
A:Accession: A42771
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2829 <GAL>
A:Experimental source: Belem strain, merozoites
A>Note: sequence extracted from NCBI backbone (NCBIN:108114, NCBI:P.108115)

Query Match 5.8%; Score 232; DB 2; Length 2829;
Best local Similarity 18.8%; Pred. No. 0.1;
Matches 190; Conservative 159; Mismatches 315; Indels 346; Gaps 46;

QY 39 DVGMMHKEEKKNDENK-----KDEERKKIQEHLKEIMKHIVK----- 79
Db 706 DINALLEEKFEVTENKSTLEMLKDEE---MEEKLDAKETFAKLNFVSDKLTDTVYT 761
QY 80 ----EYKGEAVKKEAEKLEKVPDLYMYKAIGKI-----YIVGCD--ITKHI 125
Db 762 KMSAEVTNMEGIIKKEIAQKQFENVHKKMKFSDAFSTKFEALQNSMQQYNOEGALIKKH 821
QY 126 SLEALSDEK---KKIKDIYKQALLLHEHYVVAKEGT---EPVLVIQSSSEDVVENTEK 176
Db 822 QNRSEKEEEYFKNESVEEDLSRETEOEYTKHKHNNFSRRKGEISAITJNMREYINKIES 881
QY 177 ALNYYVEI-----GKILSRDILSKINQPYOKRLDVLNFKNA--- 213
Db 882 QLNYYGYIEKFEYSLIGDQNEVSTAKALKETIVSDLSLKDIDQYETERKEKTSAVENTVST 941
QY 214 --SDSDGODLL-----FTNOLKEHPTDES----- 235
Db 942 IQSLSKAIDSLKLKLNINSCKKYNDIDILRSKITLREEVOKEMPKRGDKGENTTAL 1001
QY 236 -VEFLLEQNSVEQVAFKAFAYIIEPOHNDVLOLYAPFAFNWMDKFNEOEINLSLEELKD 294
Db 1002 LKSLRDKMGKINELKNDGRSLNLDTFKKEDLLKTFYSE-----SKSKIHLSKDGKP 1052
```

```
QY 295 QRMLSREKMEKTIQO-----HYQHWSDLSSEGRGLLKKIQI-----PIEPKKDIIHS 343
Db 1053 QDPLNRIDEMWDIKRVDYDELNVNQ-----VISEKVTLPFNKNSVYIEAMSHINTVAHG 1108
QY 344 LSOEKKELLRIOQDSDFLSTEEKEFLKL-----QID--INDSLEEKELNLR--- 392
Db 1109 ITSNNKILKSVKKEVEKLNINVEQNEDEYKKYKNPENKOLEAIGSMS-KLKEVINKHVS 1167
QY 393 --IQVDS--SNPLSEKER-----EFLKKYLKIDQPYDINQRLD 427
Db 1168 EMTOLESTANTLKSNAGKENEHDLFEELNKTGQMRDIYEKLKIALELKEGVANE-LKD 1226
QY 428 T-----GGIDPSINLD-----VRKQYKDIQNI-----DAL 455
Db 1227 ANEKANKVEPEPERNIIGHVLEIRITVEKDKAGKVEEMNSLTKIEKLTQEDSDSQNEL 1286
QY 456 LHOISGSTLYN-KIY-----LYENMNINLTATLGA 485
Db 1287 VTTISITHLENAKYEDVIKRNEDSIOLEKAKSLETLDKMKLYVOOVNMLQSAIOG- 1345
QY 486 DLYDSTNTKINRGI---FNEFKNFKYSISSNMYIVDINERPALDNERLKW-RIQLSPD 541
Db 1346 -----NAGISKELNELKGVIELLISTNYSI-----LEVYKKNSSSES 1382
QY 542 TRAGYLENGKIL-----QRNIGLEIKDVQIIOQ---SEKEYIRIDAKVVPKSKID----- 589
Db 1383 VRESQLANGETTKAGEGKKNASARLAERKEKEDIVKDLVDSDIDVV---KKEGIIKRE 1439
QY 590 -TKIOEAQLNINDEMNKALGLPKYTKLTFEVN-----RYASNIVESAVYILNEMK 640
Db 1440 ILKKKESALTF--WEES---EKFKQKSSIMENAKGKKKIEFLKNNGDG-----K 1486
QY 641 NNIOQSLIKKVTNYLVLDNGRPFVFTDITL-----PNIAEQYTHODEIYQVHSGLYV 693
Db 1487 ANITDSOMEVEGVNYSKAEHAFHTEVAQVDTKAFCSISYAVYTKMNLFNESIMKEVKY 1546
QY 694 PESISILLHGPSKGVLELRNDESGFIHEFGHAVDYAGYL-----LD 734
Db 1547 K-----CEKKND-----AEKYSALKPYDGRITARVSENERKISSELKAKVE 1590
QY 735 KNOS----DLYTNSKKRFDIFKEE--GSNLSYGRF--NEAEFFFAEFLRM 777
Db 1591 KKESSQLDNVSTKSLDIDNCRQOLDVLSIGVKNQNALQYFPSAKSM 1640

RESULT 11
S28061
SCP1 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 08-Oct-1999
C:Accession: S28061
R:Meuwissen, R.L.J.; Offenberg, H.H.; Dietrich, A.J.J.; Riesewijk, A.; van Iersel, M.
EMBO J. 11, 5091-5100, 1992
A:Title: A coiled-coil related protein specific for synapsed regions of meiotic proph
A:Reference number: S28061; MUID:93098884
A:Accession: S28061
A:Molecule type: mRNA
A:Residues: 1-946 <MEU>
A:Cross-references: EMBL:X67805; NID:957212; PIDN:CAA48006.1; PID:957213
C:Genetics:
A:Gene: SCP1
C:Keywords: DNA binding

Query Match 5.8%; Score 229.5; DB 2; Length 946;
Best local Similarity 21.8%; Pred. No. 0.032;
Matches 192; Conservative 154; Mismatches 331; Indels 205; Gaps 48;

QY 44 VKEKKKKDENKKRDEERKNTQEEHLKEIMKHIVKIE-----VKGEAVK----- 88
Db 77 LKQKENLQENRRIETIAQRAIQELQFENKRVSLKEEIEIQENKDLIKENNATRRHMCNLL 136
```

Qy	89	KEAAEKLKPSVLEPMYKAIGKIYI	-VDGOTKEHI	-----	SLKLSDDK	-----	XIK	138			
Db	137	KETCARSAEK	-TSKYEYEEETROYVVDLNNIEKMI	LA	FELKRAQAA	NALNEHFKL	194				
Qy	139	DIYKDALLHEHY	----	YAAEGYEPVLVIOSSEDEV	ENTKALNYYE	IGKILSRDLSK	195				
Db	195	EDHEKIHLEEEYOKERYVNNKENOVSL	ILIOSTER	-----	ENMKKIDTFLLE	-----	ESRD	-K 245			
Qy	196	INOYOKF	-----	LDVLNITKANSDDODLFTNO	-----	LKEH	-----	PTDSYVEF 238			
Db	246	ANOEEETKLADENMLKELNEKDH	L	SELEIDKMSMORSKSTOKTLEB	LOATATYQOL	305					
Qy	239	LEONSNVEQEV	-----	FAKAFAYIE	-----	POHRDVLQIAYAE	-----	272			
Db	306	TEEKAOQMEELNKATHTSLVLT	ELKATCTOTBELLRTEQOR	LENNEQOLKITMELQK	365						
Qy	273	--AANYMDKF	NEOEINLSLEELK	-----	DORULSKREWEKIKOHYQWSOSLS	--	EE	321			
Db	366	SSELEEMTKFKNNKVE	----	ELEEBETILAEQKLDKOV	EKAELEQKBOELFTLQOT	423					
Qy	322	GRGLKLOTPREPKKODIHSLOE	-----	EKELKRIOD	-SSDPTSEKEF	---	370				
Db	424	REKEIHLEQVYVTKISEEHYLYKQV	BEKTELEKRLNI	ELTANSMDLLENKKLYOE	483						
Qy	371	-----	LKKLQIDIRDSLSEERK	-----	ELNRIQDSSNPLSEKEKBFK	-----	KLK	413			
Db	484	ASDVAYLELTKHQEODI	INCKQOEERMYLQJETTELEK	EMNLRDELESYRKFETIQGD	EVAK	543					
Qy	414	LDIO	-----	PVINO	RLODTGG	LDPISINDVAKQYKRO	QND	DALLHO	-----	SIG	461
Db	544	LDKSEENARSTIEYELK	-EKOKMITLEKNCNN	----	KKOLENKSNNLE	-LHENKALKK	599				
Qy	462	STLVNK	-IYLEYMNNINNLATGLADLV	SDTWTIKIN	-GJIEFEFKNFY	-SISSNYMI	518				
Db	600	SSAENKQLMNVE	-IKVAKLELELA	-----	STKQKPEEMINNYOKELEIKKISEKLL	650					
Qy	519	VDINERPALDNERLKWRIQ	LSPTPRAGY	-----	LENGKLILO	RIGLEIKDVOIITKO	570				
Db	651	GEVEKAKATYDEAVK	-LOKEIDLRCOHKIAE	VAVALMEKHKHQYD	RIVEERDSEGLAYKN	708					
Qy	571	SEKEYIRIDAKVPKSKID	KIOE	-----	AOJINONEKMA	GLL	-PKYKLTTPVNHRY	624			
Db	709	REOE	-OSSAKVAALETELSIR	NELVLSLKKQLEVEKEE	KKLMEOENAI	LLTDKKDKKI	766				
Qy	625	ASNIYESAVYLLNEMK	----	NNIOSDLIKKVTNLYVANG	-----	REYFTDITPLINAE	674				
Db	767	QASLSEPE	-ATSMKRDSTTPSONIS	RSSMSOGSKKDND	SLRAKAKSLSTYTK	824					
Qy	675	QYTHODELYEOVHSGLYVES	RSILHPSK	-----	GVELRDNSE	-----	GFITHERG	722			
Db	825	EYT	-----	VTPTKKSTIYORENKRYLPTG	GSNNKRRKATVEFEDV	SDSSSETTDL	SLYSE	--	877		
Qy	723	HAYDVAGYLLDKNOSD	-IVTNSK	KFIDIFKREGS	NLSYSG	762					
Db	878	----	EDISNRIYNNTPDSHLV	TPKQOTPLISTP	ASPSTKfG	916					

RESULT 12
S49461
snapltonemal complex protein 1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-Nov-1999
C:Accession: S49461: S59599
R:Julien, S.; Luc, M.; Francois, C.
Submitted to the EMBL Data Library, October 1994
A:Description: Cloning and sequencing of the murine SCP1 cDNA.
A:Reference number: S49461
A:Accession: S49461
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-993 <JUL>
A:Cross-references: EMBL:Z38118; NID:g1360015; PIDN:CAAB6262.1; PID:g55860303
R:Sege, J.; Martin, L.; Cuzin, F.; Rassoulzadegan, M.

Biochim. Biophys. Acta 1263, 258-260, 1995
A>Title: cDNA sequence of the murine synaptonemal complex protein 1 (SCP1).
A/Reference number: S59599, M0ID:96004899
A/Accession: S59599
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-993 <SAG>
A/Cross-references: EMBL:Z38118; NID:q1360015; PIDN:CA86262.1; PID:g55860303

Query Match	5.7%;	Score 228.5;	DB 2;	Length 993;
Best Local Similarity	22.0%;	Pred. No. 0.038;		
Matches 202;	Conservative 144;	Mismatches 336;	Indels 237;	Gaps 49;

QY	44	VKEEKKMKDEKRRDERNNKTOEENHLEIKHNHJYKLEVEKEEVAUK	88
Db	124	LKQENKMLQENKRIIEAQRKA---IQELOFENKYSKLAEEDIQENKDLKENNATIM	179
QY	89	-----KEAAEKLLEKVPSDVLEMTAKIGCKYIT--VDQDTKH-----SLEASDDKA----	135
Db	180	CNLLKCEKARSAAK--TNKYEYEEETROYVLDNSIEMKILAFELBRLQVEMARLEMH	237
QY	136	-KIKDITGKDALHEEH---VYAKEGEYPLVIOSSDYENETKALNYUEYLGKILSRD	191
Db	238	FKLEDEHEKIOHLEEEYQKEVKNKENYSELLQISAQ--ENKMKDITELLE-----ESRD	291
QY	192	ILSKINOPYOKF-----LDVLTITKNASDSGDOL-----LFTNQLKEHPTDSVEF	238
Db	292	---KANOLEEYTKLODENLKELSKKKHLTSELEDIMKSQMSSTOKALEEDQIATKTT	348
QY	239	LEQNSNEVOE-----VFAKARAYITE-----PQNHVLOLAPR	271
Db	349	ISO--LTEREKQOMELKKAATTISFVUTTELKATTTCTBELLTETEOORLENEBOLKITIV	407
QY	272	E-----AFNYMOKF--NEOEINLSLELK-----DOQMLSRYEKMEKIKONYOAMSDSL	319
Db	408	ELQKSNBLEEMTQFKNNKEV--ELEELKNTILADQKILDEKQVKEGLAELOEKEQOELT	465
QY	320	---EEGGLKKTLOIPLEPKKDILHSLO-----EKELLKRIQIDSS--DFLSTEEK	368
Db	466	FLITTREREYHDLQOYTATVTSEONHLYKQVEEMKTELEKEKTLNPELTAQCDMLLEAK	525
QY	369	EF-----LKKLIDIRDSISEBEKELYNRIQVDSNPJSEKE-----KE	407
Db	526	KFYQASDMALTELKHOEDILINCKQOER--ILKQIE-----NLEKEMHRLDESVRKE	579
QY	408	FLK-----KKLDIOPYDINO-----RIODTGLDPSINLDYVKQYKRIQNTIDALL	456
Db	580	FIOQGEVCKLKDSEENARSISECEVLKKEKOKMILSEKCNL--KKQVENKSKIEE--L	636
QY	457	HOSIGSLYLNK-----ILYEMNNINLTALGLADVDSNTQNTKINGINFEFEKKNK	509
Db	637	HOE--NKLTKKSSAEIQOLANYE--IKVSKLELE-----ESTKOR--PEEMTNNO	683
QY	510	-----YAISSNYMVDINERPALDNERLWKRIOLSPDTRAGY-----LENGKLIILOR	556
Db	684	KEIENKRISEGLKILGEYERAKATGDAVUK--LQKEIDLRCQHKITAEVVALMEKHNQYOK	741
QY	557	NIGLEIDVOQIKOSEKEFYIRIDAKVYKSKIDTKIOE-----AQINDQENMKALGLEK	611
Db	742	IVEERDSELGLYKRRDE--OSSAKLITELTISINRELNYSILKOLEIEKEEKEKLMKAK	799
QY	612	-YTKLITFENYHNRASNIVESAYLILNEMK---NNIOSDLITKVTYVLYVDG-----NG	660
Db	800	ENTYAILDKDKDKIQAASLESPE--AISMKFPDSKTPPSQMSIKLSSMSQSKSKDRDML	857
QY	661	REVFETDITLPNIAOYTHODEIYEQVHVK--GLVYBESRSTILHGPSKGYELARNDSE---	715
Db	858	RASAKSILPTVTEYKEYVTAKPTKKSIIYQENBKYIIPGSSNKKRKTAFEEVDNSDSEAD	917
QY	716	--GIHFEGHADVYAGYLLDKQNSD---LVYMSK-----KFIDLFKEGSGN	757
Db	918	LLSLVSE---EVSNSRLVDNPNRPSDHLILVTKPKQPTSLSTPASPMSFGSLKKKREDR	972


```

OY      758  LTSVCRTNEAEFFFAEFL  776
      | : : : | | : |
Db      973  WTTIAKIDKKRLKEAEKL  991

RESULT  13
B1603
RESA-H3 antigen PF80915W - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: B71603
R:Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
: Partea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Ref: accession number: A71600; MUID:99021743
A:Accession: B71603
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1558 <GAR>
A:Cross-references: GB:AE001424; GB:AE001362; NID:g3845307; PIDN:AACT1972.1; PID:g3845307
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PF80915W

```

```

Query March 5.64: Score 224.5; DB 2; Length 1558;
Best Local Similarity 18.54; Pred. No. 0.096;
Matches 156; Conservative 123; Mismatches 260; Indels 305; Gaps 29;

OY 44 VKEKKNNDENRKRKDEENKTKTOEHLKEIMKHIVIEVKGEEAVKKAEEKLLEKVV----- 99
      :::::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 659 IEKKLEELHENVLSALENTOSEEKEKEVIDY-----EEVKEVATLIIIEVEQAE 710
OY 100 ---PSDVLKMKAIKGKYYIVDGDITKHI-----SEALSE 132
      -  :::::  :::::  :::::  :::::  ||
Db 711 EESASTIIEIFENLEENAVESNENVAENMLEKINETVFMTVLDKVEETVEISGESIENEM 770
OY 133 DKKKIKDLYGDALLHEHYVYAK -EGYEPVLVITOSSE-----DVENFEKAL 178
      :::::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 771 DKAFSEIFDNVKGQIENLITGMFRSIETSIYIOSEEVKVDLNNVYSSILNDIENMKEG 830
OY 179 -----NYYEIKGLSRD-----ILSKINQYOKFLVDLYNIRK 211
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 831 LNKLENIISTEGVQETVEHVEQVNYVDVAPAMKDQFLGILNEAGGKEMFNEVDYFK 890
OY 212 NASDSDGDDLEFTNOLKEHPIDFSVEFLQNSNENVEQVEFAKAFATYIEPQHRDVIQLYAP 271
      :::::  :::::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 891 SES-----DVITVEIKQEP-----VQKEVEKTVSIIIEEMENIVDVEE 931
OY 272 EAFNYMDF-----NQEINLSLEELDQMLSKYKWEKIR 308
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 932 EKEDLTDMKIDAVEISIEISDSKETESIKDKKDVSLVVEEVDNDMDSEVKEVLEIK 991
OY 309 OHYQHM-----SDSLSEGRGL-----LKLIQIPLE-----PKKD 338
Db 992 NMEELMKDAVEINDITSKLIETQELNVEADLIKDMKLEKLEKALSSESKETIIDKD 1051
OY 339 DIHLSQBEKEL-----LKRIQIDS-----SDFLSTEB-----KEF----- 370
      :::::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1052 DTLEVEIEEHDITTLDEVELKQVEEDKIEKVSDLMDLDEEDILKEVKEIKELSETELE 1111
OY 371 ----LKLIQIDRSLSEBEKELNRIQVDSNPLSEKKEKELKLIKIDIQPYDINQRLQ 426
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1112 DYKELKTIEITDLEKKEIKEDKHEKF-----EEBAEIKLEADI-----LK 1154
OY 427 DTGGLIDSPSINLDVRYOKYRDQINI-----DAL-----LQSIGSTLYNKIY 469
      :::::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1155 EVSSLEVEBEKKLEVUHELKEVEHIIISGDAHIKGLDEDDLEEVYDLDGSLIDMLKGOME 1214
OY 470 L--YENMINNLTATLGLADVLDSNTKINRGIFNEFKFNKRYTSSSNMIVDINERPAL 527
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

```

```
Dd      1215 LGDDMDKSLBVTAKKGERRERLSLDVLSALMGDDEQMTRRKA-----QRKL 1263
Oy      528 DNERLKMRIQSPDTRAGYLENGSKLLÖRNIGLEIKD-----          564
Db      1264 ECVLLKEVEEKP-----KITKKKVREDINOKESPKEIYEVENKDEDIDEE 1314
Oy      565 -----VOIIOSEKEYIRIDAKVVPKSIDPTK 591
Db      1315 DYEDIEDBKVEDIDEDIDEDIDEGDKDEVITDLIVKERIREKVEK---KKLEKK 1371
Oy      592 IOEQALNINQEWNKALGLPYRTKLITFNHNHRASYNIVESAYALLINEMKNNIQSOLIKV 651
Db      1372 VEEGVSGLKRRHDEVMA---KYVQIRKEV--DKREVSKALESKNDVTNVILKÖN--QQFESV 1425
Oy      652 TNYL 655
Db      1426 KNFV 1429
```

RESULT 14
T18427
hypothetical protein C0335C - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence
C:Accession: T18427
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18427
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-3724 <LAW>
A:Cross-references: EMBL:Z698547; NID:el325376; PID:el325379; PIDN:CA011104.1
C:Genetics:
A:Introns: 307/1, 1545/2
A:Note: C0335C

```

Query Match 5.6%; Score 221.5; DB 2; Length 3724;
Best Local Similarity 19.5%; Pred. No. 0.38;
Matches 170; Conservative 174; Mismatches 321; Indels 207; Gaps 39.

OY 45 KEREKNDENKKRDEKRNKTOEHLKEIKMHVIE-----VKGE 84
      ||: ||| : : : ||: ||: |||
DB 497 KJLOERKDEKRNKNDTYINNNEEQFDLNRINKIESINNNDNNNNINNKKEFNKIRE 556
      ||: ||| : : : ||: ||: |||
OY 85 EAVKKEAEKLEKXPS---DYLEWKA-----IGKIIYVDDGTR- 123
      ||: ||| : : : ||: ||: |||
DB 557 HILNEKSIKHLGSDPSRDKETKLTYYNKNEDSTFELKELEITNNKVNVEEDIGS 616
      ||: ||| : : : ||: ||: |||
OY 124 -----HISLEALSDE-----KKRKIDYKDALHHEYYAAEGEPPVLIOSSEDY 170
      ||: ||| : : : ||: ||: |||
DB 617 NEDDEYIHLKENLKEDANEYNNDEKNNKNTKEILKSNYLENERKRTLEELKLRGNKI 676
      ||: ||| : : : ||: ||: |||
OY 171 VENTERKALNVYEIGKIISROL-----SKINQ-----PYQFLDY---LNTIKASDSD 217
      ||: ||| : : : ||: ||: |||
DB 677 FKRDK---YNSLEVIINEIQINEENKINIDOGNISKOKIIOSSSFNTOTFNKIDS 732
      ||: ||| : : : ||: ||: |||
OY 218 GODLFFTNOLKEHPIDFSVEPLEONSNEQVEFAFAFYIIPQHRDVLQTLAPFAFNM 277
      ||: ||| : : : ||: ||: |||
DB 733 LMDDELEKERRKKSQHFIDNLVKADKNELSEINIKICOMNINNIYDESINNIIYDES INNII 792
      ||: ||| : : : ||: ||: |||
OY 278 DKFNEOEINL---SLEELKDQRMLSRYEKWEKIKOHYQMSLSLSEEGGLLKKIOPIE 334
      ||: ||| : : : ||: ||: |||
DB 793 --YDESINNIIYDESINNIIYDENINNIIYD--ENINNIIYDENINNIIYDEGINKICDDNI -LE 847
      ||: ||| : : : ||: ||: |||
OY 335 PKK-----DRII-----HLSOEKELKRIQIOSSD-----FLSTEEKFELKKIOP 375
      ||: ||| : : : ||: ||: |||
DB 848 KKNITKTNDIYQVEENNEIEKNEELMISLNKINNTYNNKFNENVDIFIKIKRESLKLKD 907
      ||: ||| : : : ||: ||: |||
OY 376 IDIRDSLSEEEKELLNRIQV-----SSNPLS-EKEEFLKLKLDIOPYDIN 422
      ||: ||| : : : ||: ||: |||
DB 908 KNIKONNNDDEYIMONFENDEIINHKMEITNKEDLPLEINTQNERIENLDJRRKKYND 967
      ||: ||| : : : ||: ||: |||

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:51:33 ; Search time 44.8 Seconds
(without alignments)
636.724 Million cell updates/sec

Title: US-09-747-521-2_COPY_1_778
Perfect score: 3987
Sequence: 1 MNKEFKIVKISMCLVTAL.....TSYGRTNAEFFAEFRIMH 778

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3987	100.0	809	LEF_BACAN	P15917 bacillus an
2	501	12.6	800	CYAA_BACAN	P40136 bacillus an
3	232	2869	1	RBP1_PLAVB	000798 plasmodium
4	229.5	5.8	997	SCP1_RAT	003410 rattus norv
5	228.5	5.7	993	SCP1_MOUSE	062209 mus musculu
6	223	5.6	794	HMR_MOUSE	000547 mus musculu
7	219.5	5.5	2663	CENE_HUMAN	002224 homo sapien
8	218.5	5.5	1075	Y124_METJA	057588 methanococ
9	215	5.4	2245	YD86_SCHPO	P54697 dictyostell
10	209.5	5.3	1957	YD86_SCHPO	010411 schizosacch
11	209	5.2	1679	Y109_YEAST	P40457 saccharomyc
12	208.5	5.2	691	Y104_YEAST	P40460 saccharomyc
13	208	5.2	1330	SMG3_YEAST	P47037 saccharomyc
14	205.5	5.2	756	Y348_MYCGE	Q49419 mycoplasma
15	205.5	5.2	1875	MLP1_YEAST	002455 saccharomyc
16	202	5.1	2710	TOXA_CLODI	P16154 clostridium
17	202	5.1	2748	NUM1_YEAST	000402 saccharomyc
18	201.5	5.1	3911	AKA9_HUMAN	Q99996 h a kinase
19	201	5.0	1790	USO1_YEAST	P25386 saccharomyc
20	200	5.0	845	SCP1_MESAU	060565 mesocricetu
21	199.5	5.0	1251	RBP2_PLAVB	000799 plasmodium
22	199.5	5.0	2116	MYG2_DICDI	P08799 dictyostell
23	199	5.0	1312	RA50_YEAST	P12753 saccharomyc
24	197.5	5.0	1427	REST_HUMAN	P30822 homo sapien
25	197.5	5.0	1928	MYG1_YEAST	P08664 saccharomyc
26	197	4.9	1940	MYH3_RAT	P12847 rattus norv
27	196	4.9	976	SCP1_HUMAN	Q15331 homo sapien
28	196	4.9	1939	MYH6_MESAU	P13339 mesocricetu
29	195.5	4.9	1805	HMR_MYCGE	P47460 mycoplasma
30	194.5	4.9	3210	CENP_HUMAN	P49454 homo sapien
31	194	4.9	1727	ALM1_SCHPO	Q9UKS5 schizosacch
32	192	4.8	2230	GOG4_HUMAN	Q13439 homo sapien
33	191.5	4.8	944	NDP1_YEAST	P32380 saccharomyc

ALIGNMENTS

```

RESULT 1
LEF_BACAN          STANDARD;          PRT;          809 AA.
AC      P15917;
DT      01-APR-1990 (Rel. 14, Created)
DT      01-APR-1990 (Rel. 14, Last sequence update)
DT      15-JUL-1998 (Rel. 36, Last annotation update)
DE      LETHAL FACTOR PRECURSOR (EC 3.4.24.-) (Lef).
GN      LEF.
OS      Bacillus anthracis.
OG      Plasmid pXOI.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group;
OC      Bacillus/Staphylococcus group; Bacillus.
OX      NCBI_TaxID=1392;
RN      [1]
RP      SEQUENCE FROM N.A. AND SEQUENCE OF 34-49.
RX      MEDLINE=90034185; PubMed=2509294;
RA      Braag T.S., Robertson D.L.;
RT      Nucleotide sequence and analysis of the lethal factor gene (lef)
RT      from Bacillus anthracis." ;
RL      Gene 81:45-54(1989).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Lowe J.;
RL      Submitted (APR-1990) to the EMBL/GenBank/DBJ databases.
RN      [3]
RX      MEDLINE=95154669; PubMed=7851740;
RA      Kochl S.K., Schlavo G., Mock M., Montecucco C.;
RT      "Zinc content of the Bacillus anthracis lethal factor." ;
RL      FEMS Microbiol. Lett. 124:343-348(1994).
CC      -1- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,
CC      AGENT WHICH INJECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE
CC      DEATH. LEF IS THOUGHT TO BE A LETHAL FACTOR THAT, WHEN ASSOCIATED
CC      WITH PA, CAUSES DEATH. LEF IS NOT TOXIC BY ITSELF. PA IS THOUGHT TO
CC      BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC CELLS, THEREBY
CC      FACILITATING THE INTERNALIZATION OF LEF OR EF.
CC      -1- SUBUNIT: SECRETED ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT
CC      PROTEINS. A PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LEF) AND AN
CC      EDEMA FACTOR (EF). NONE OF THESE IS TOXIC BY ITSELF.
CC      -1- SUBCELLULAR LOCATION: SECRETED.
CC      -1- SIMILARITY: THE PA-BINDING REGION IS FOUND IN BOTH B.ANTHRACIS EF
CC      AND LEF.
CC      -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M34 (ZINC
CC      METALLOPROTEASES).
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: M29081; AAA9216.1; -
CC      DR      EMBL: M30210; AAA22569.1; -

```

```
DR PIR: JQ0032, JQ0032.
DR MEROPS: M34..001; Anthrax_toxinA.
DR InterPro: IPR0003541; Zn_MTPepidase.
DR InterPro: IPR000130; Zn_MTPepidase.
DR PROSITE: PS00142; ZINC_PROTEASE_1.
KM Hydrolyase; Metalloprotease; Zinc; Toxin; Signal; Repeat; Plasmid.
FT SIGNAL 1 33
FT CHAIN 34 809 LETHAL FACTOR.
FT DOMAIN 34 293 PA-BINDING REGION (POTENTIAL).
FT METAL 300 420 REPEATS.
FT METAL 719 719 ZINC (CATALYTIC) (POTENTIAL).
FT ACT_SITE 720 720 POTENTIAL.
FT METAL 723 723 ZINC (CATALYTIC) (POTENTIAL).
SQ SEQUENCE 809 AA: 93786 MW: 8C16B4D727310AB CRC64;

Query Match 100.0%; Score 3987; DB 1; Length 809;
Best Local Similarity 100.0%; Pred. No. 5,2e-148;
Matches 778; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKKFKIVISMGLVTAITLSCGPVPIPLVQAGHGDMHVKKEKKNDEKRNKDEE 60
D 1 MNKKFKIVISMGLVTAITLSCGPVPIPLVQAGHGDMHVKKEKKNDEKRNKDEE 60
QY 61 RNKQOEHLKELMKHYIEVKEGEAVKKEAEKLEVPDVLDMYKAIGKTIYVGD 120
D 61 RNKQOEHLKELMKHYIEVKEGEAVKKEAEKLEVPDVLDMYKAIGKTIYVGD 120
QY 121 ITKHISLESEDKKIDYIGKDALHEHYVYAKEGEPVLVIOSSDYENFEKALNV 180
D 121 ITKHISLESEDKKIDYIGKDALHEHYVYAKEGEPVLVIOSSDYENFEKALNV 180
QY 181 YYEIGKILSRDILSKINPYOKFLDVLNTIKNASDSDGDLFTNQLKHPDSEVFLE 240
D 181 YYEIGKILSRDILSKINPYOKFLDVLNTIKNASDSDGDLFTNQLKHPDSEVFLE 240
QY 241 QNSNVOVFAKAFAYIEPQHRVLYLAPAEFVYMKFNEQELNLSLELKQRMISR 300
D 241 QNSNVOVFAKAFAYIEPQHRVLYLAPAEFVYMKFNEQELNLSLELKQRMISR 300
QY 301 YEKKEKIKQHHMSDSISEGRCGLKQLPIEPKKDDIHSLSQSEKELLRIOISS 360
D 301 YEKKEKIKQHHMSDSISEGRCGLKQLPIEPKKDDIHSLSQSEKELLRIOISS 360
QY 361 DFLSTEKEFLKQLDIDRSLSEEEKELNRIOVDSNPLSEKKEFLKLKLDIOYPD 420
D 361 DFLSTEKEFLKQLDIDRSLSEEEKELNRIOVDSNPLSEKKEFLKLKLDIOYPD 420
QY 421 INORLODTGGLIDSPINLDVRKQYKRDIONIDALHOSISTLYNKLYLENMNINLT 480
D 421 INORLODTGGLIDSPINLDVRKQYKRDIONIDALHOSISTLYNKLYLENMNINLT 480
QY 481 ATLGADLVSDNTKINNGIFNEFEKNKFKYSSNMYIVDINERPALNRLKRIQLSP 540
D 481 ATLGADLVSDNTKINNGIFNEFEKNKFKYSSNMYIVDINERPALNRLKRIQLSP 540
QY 541 DTRAGYLENGKLLIQRNIGLEIKDVQIKQSEKEYIRIDAVVPKSKIDTKIQEAQLIN 600
D 541 DTRAGYLENGKLLIQRNIGLEIKDVQIKQSEKEYIRIDAVVPKSKIDTKIQEAQLIN 600
QY 601 QEMKKAIGLSPYTKLTFNNVNRASNVESAYILLNEMKNNOISDLLKKTATNYLYDNG 660
D 601 QEMKKAIGLSPYTKLTFNNVNRASNVESAYILLNEMKNNOISDLLKKTATNYLYDNG 660
QY 661 RFVFTDITLPIAEOYTHODEIEQVHSGLYVPESSRIILHGPCKGYELRNDSCGFHE 720
D 661 RFVFTDITLPIAEOYTHODEIEQVHSGLYVPESSRIILHGPCKGYELRNDSCGFHE 720
QY 721 FGAHVDDVAGYLLDKNOSDLVTSNKKFLIDFKREGSNLTSYGRNTAEFFAEAFRLMH 778
D 721 FGAHVDDVAGYLLDKNOSDLVTSNKKFLIDFKREGSNLTSYGRNTAEFFAEAFRLMH 778
```

```
RESULT 2
ID CYAA_BACAN STANDARD; PRT; 800 AA.
AC P40136;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CALMODULIN-SENSITIVE ADENYLATE CYCLASE PRECURSOR (EC 4.6.1.1) (ATP
DE PYROPHOSPHATE-LYASE) (ADENYLYL CYCLASE) (EDEMA FACTOR) (EF).
GN CYA.
OS Bacillus anthracis.
OC Plasmid pXOI.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OC NCBI_TaxID=1392;
RP [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89138004; PubMed=2906312.
RA Escuyer V., Duflot E., Sezer O., Danchin A., Mock M.;
RT "Structural homology between virulence-associated bacterial adenylate
RT cyclases."
RL Gene 71:293-298(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89211974; PubMed=3149607;
RA Robertson D.L., Tippetts M.T., Leppia S.H.;
RT "Nucleotide sequence of the Bacillus anthracis edema factor gene
RL (cya): a calmodulin-dependent adenylate cyclase."
RN [3]
RP SEQUENCE FROM N.A.
RA Escuyer V., Duflot E., Mock M., Danchin A.;
RT "Nucleotide sequences expressing adenylate cyclase from B.anthraxis,
RT proteins having the activity of this adenylate cyclase and biological
RT uses."
RL Patent number EP0366550, 02-MAY-1990.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=88198021; PubMed=2834337;
RA Tippetts M.T., Robertson D.L.;
RT "Molecular cloning and expression of the Bacillus anthracis edema
RT factor toxin gene: a calmodulin-dependent adenylate cyclase."
RL J. Bacteriol. 170:2263-2266(1988).
RN [5]
RP SEQUENCE OF 34-48.
RX MEDLINE=89211974; PubMed=3149607;
RA Schmidt J.;
RT Unpublished results, cited by:
RL Robertson D.L., Tippetts M.T., Leppia S.H.;
RL Gene 73:363-371(1988).
RN [6]
RP MEDLINE=93119764; PubMed=8418825;
RA Danchin A.;
RT "Phylogeny of adenylyl cyclases."
RT Adv. Second messenger Phosphoprotein Res. 27:109-162(1993).
CC -1- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,
CC AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE
CC DEATH. EF IS A CALMODULIN-DEPENDENT ADENYLYL CYCLASE THAT, WHEN
CC ASSOCIATED WITH PA, CAUSES EDEMA. EF IS NOT TOXIC BY ITSELF. PA IS
CC THOUGHT TO BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC CELLS,
CC THEREBY FACILITATING THE INTERNALIZATION OF EF OR LF.
CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC -1- SUBUNIT: ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT PROTEINS, A
CC PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN EDEMA FACTOR
CC (EF). NONE OF THESE IS TOXIC BY ITSELF.
CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-2 FAMILY.
CC AND LF.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M23179; AAA22374.1; -
DR EMBL: M24074; AAA79215.1; -
DR EMBL: A07289; CA00652.1; ALT_SEQ.
DR Interpro: IPR003541; Anthrax_toxinA.
KW Lyase; CAMP synthetase; Toxin; ATP-binding; Signal; Plasmid.
FT SIGNAL 1 33
FT CHAIN 34 800 CALMODULIN-SENSITIVE ADENYLATE CYCLASE.
FT DOMAIN 34 288 PA-BINDING REGION (POTENTIAL).
FT NP_BIND 289 680 CATALYTIC.
FT NP_BIND 347 354 ATP (POTENTIAL).
FT CONFLICT 350 350 V -> E (IN REF. 2).
FT CONFLICT 510 510 Q -> T (IN REF. 2).
FT CONFLICT 512 513 EW -> RM (IN REF. 2).
FT CONFLICT 760 760 V -> L (IN REF. 3).
SQ SEQUENCE 800 AA: 92477 MW: 94758456 DFC5A6 CRC64;

Query Match 12.6%; Score 501; DB 1; Length 800;
Best Local Similarity 24.6%; Pred. No. 5.6e-13;
Matches 202; Conservative 145; Mismatches 281; Indels 194; Gaps 36;

OY 4 KKEFT---KVISMSCLVATITLSCVPFIPVLOGAGHGDCVGMHVKKEKNKDKNRKDE 59
DB 3 RNFETPNKFSIISFVLLRAISSQAEVNA-----MNEHTESDIKRNHKT 49
OY 60 ENRKOEHLEKMKHIVIEVKGEEAVKKEAEKLEVPDSVLYMKAIGIKIYVG 119
DB 50 ENRKEEKEFKOSINLVTEFTNELDKIQOTDILKIKPQVLEIYSELGSEIYFTDI 109
OY 120 DITKHSLEALSEDKKKIKDIYKDALLHEHYVYAKEGEYEVLYIOSEEDYVENTEKALN 179
DB 110 DLVEHKELODSEEEKSNMNSRGEKVPASRFVEKKREPKLII-NIKQYAINSSQSK 168
OY 180 VYVEIGKILSRILSKINPYOKFLD-VLNTIKNNS-DSGODLLFTNOLKE---HPT 232
DB 169 VYVEIGKIGLIDILSK-----DKSIDPEFLNLKSLSDSDSLDFSOFKFEKLELNK 223
OY 233 DRSVEFLEQNSNEVOEFAKAFAYIEPOHBDVLOLYAPEAFVYMDKPNQENLSLEL 292
DB 224 SIDINFIKENLFEQAFSLAFSYFAPDRKRYLEIYAPDMFYMKL----- 271
OY 293 KDQRLSRYEK-WEKIKOHYOHMSDSLSEEGRLKKLQIPLEPKDDIHSLSOEKE 350
DB 272 -----EKGGFEKI-----SESLKKEG-----VEKDRIDVL---KGEKA 301
OY 351 LKRIQIDSSDPLSTEKEFLKKLOIDIRDSLSEKELNKRQVSSNPLSEKEFLK 410
DB 302 L-----KASGIVPEHAAPFKRI-----ARELNTYLLFRPVNLAATN---LI 339
OY 411 KKLIDIOPIQDINORLODTGGLIDSPSINLDVRYKQY-----KRDIONIDALL-HO-SIG 461
DB 340 KSGVATKGLNVHCKSSDMCPVACIYPPDOLSKKHGOQLAVEKGNLENKKSITIEHGEIG 399
OY 462 SYLYNRIYL-YENMNINNLATLGLADLVSTDTNKTINRGIFNEKKNFK-----YSISSN 515
DB 400 -----KIPKLIDLRIEELK-----ENGILLKGRKEIDMGKKYLLLESN 438
OY 516 YHIVIDINERPALDNERLAKRIQLSPTRAGYLE-NGKLLIQRNIGLEIKDQY-TIQOSEK 573
DB 439 NOVYE--RISDENNEVOYK-----TKGKTIYVLEGEKFMWNIIEVMKAVEGVLPLTA 490
OY 574 EYIRIDAKVVPKSIDKPTIOEAOINIOENKALGLPKYTK-----LITFNVNRVA 625
DB 491 DY---DLFALAPSLTEIKKQIPO---KENDKVVNTPNLSLEKQGYTNLLIKIGIERK-- 541
OY 626 SNVESAVYLILNEMKNNIOSDLIKKVTNYLVNDNGRFVFTDITLPLNIAEQYTHD--EIV 683
DB 626 SNVESAVYLILNEMKNNIOSDLIKKVTNYLVNDNGRFVFTDITLPLNIAEQYTHD--EIV 683

DB 542 ---PDSNKGTLNMQOKMDRL-----NEAVKTYGTGVDVNVNHTGEQDNDEEP 587
OY 684 EGVHSGKGLVPEBSRSTLLHG---PSKGVELRNDSEGFIEFGAVDYA---GYL----- 732
DB 588 EKDNEFIINDEEPEFLITKNWEMTGRTEKNIKNGKYLIVFNRSYNKIAFGNKAITYEMTD 647
OY 733 -LDKNOSDLVTNSKFKFDIFKEGNSNLTSGRTNEAEFFEA 773
DB 648 PITKAKINITPTSAEFT-----KNLSSIRSSNWGVYKDS 682

RESULT 3
RBP1_PLVAB
ID RBP1_PLVAB STANDARD: PRT; 2869 AA.
AC 000798;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
GN RBP1
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RX MEDLINE=92315338; PubMed=1617731;
RA Galluski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites.";
RL Cell 69:1213-1226(1992).
CC -I- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -I- SUBUNIT: HOMODIMER (POTENTIAL).
CC -I- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M88097; AAA29743.1; -
DR HSSP: P36956; IAM9.
KW Malaria; Receptor; Signal; Transmembrane.
FT SIGNAL 1 17
FT CHAIN 18 2869 POTENTIAL.
FT DOMAIN 18 2807 RETICULOCYTE BINDING PROTEIN 1.
FT TRANSMEM 2808 2826 EXTRACELLULAR.
FT DOMAIN 2827 2869 POTENTIAL.
FT SITE 1030 1032 CYTOPLASMIC.
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA: 330213 MW: 89DBE442205BECF CRC64;

Query Match 5.8%; Score 232; DB 1; Length 2869;
Best Local Similarity 18.8%; Pred. No. 0.065;
Matches 190; Conservative 159; Mismatches 315; Indels 346; Gaps 46;

OY 39 DVGAMHVEKKNKDNKR-----KDEERNKTOEHLKEIKMKHIVXI----- 79
DB 746 DINALIEVKEFVTEKNESLEMLKDEE---WEKIQDQKKEFFAKLNPVSDKLDIVYT 801
OY 80 ---EYKGEBAVKKAEKLEKLVPSDVLBYMKAIGKI-----YIVGD-ITKAI 125
DB 802 KMSAEVYTNAGIKKEIAQOFENVHKKMKFEFSAFSPKFPALQNSMQVNOGEDDAIEKKH 861
OY 126 SLEALSDEK---KKIKDIYKDALLHEHYVYAKKEG---EPVLYIOSEEDYVENTEK 176
DB 862 QNRSEKEEYFKNESVEEDLSRETEQETTKHKNNSRRKGEISAEITMMREVINKIES 921
OY 177 ALNVEYEI-----GKILSRDLISKINQYQVFLDVLNITKNA--- 213

```
Db 922 QLNITYGIEKFFSLIGDQNEVSTAKALEKTYSDLSKRDIDQYETTEREKTSAVENYVST 981
QY 214 --SDSDGDL-----FTNQLKEHPTDS----- 235
Db 982 IQSLSKAIDSLKRLNGSINNCKKYNTDIDILRSKIKTLREEVOKEMPRGDKCENTTAL 1041
QY 236 -VELEBONSNEVOEYFAKAFYIIEPOHVDYQLYAPAFYIMOKFMEQELNLSLEELKD 294
Db 1042 LKSLIRKMGKINERKLNDRNSLDTKEDDLKFESE-----SKSKIHLSDKQGP 1092
QY 295 QRMLSREKMEKIKO-----HYOHMSDSLSEGRGLKTLQI-----PIEPKDDIHS 343
Db 1093 QDPLNRIDEMEDIKRDDELAVNTQ-----VISEKVTILFKNNSTYIIEAMSHINTYAHG 1148
QY 344 LSOEKEKLRIQIDSSDFLSTEEKEFLKL-----QID-IRDSLSEEEKELNLR--- 392
Db 1149 ITSKNELKSVKEVEDEKLNVEQEDYKVKYKNEKQELAIRGMS-KKEVYINKHVS 1207
QY 393 --IOVDS-SNPLSEK-----EFLKKLKDIOPIYDINQRLQD 427
Db 1208 EMTQLESTANTLKSNAKGENEHLLEELNKTGOMRDYIEKLIKIAELKEGTVNE-LKD 1266
QY 428 T-----GGLIDSPSIND-----VRKQYKRDIONI-----DAL 455
Db 1267 ANEKANKVEPERKNIIGHVIERITVEKDKAGKVEEMNSLKTIEKLIQETSDSONEL 1326
QY 456 LHQSIGSTLYN-KIY-----LYENMINNLATLGA 485
Db 1327 VTSITKLENAKGYEDYIKRNEEDSIQIREKAKSLETFLDEMKKLVOQVNMNLSAIGG- 1385
QY 486 DLVSTDNKTKNNGI-----FNEFKNFKYSTSNMIVDINRRPALDNRLKW-RIQLSPD 541
Db 1386 -----NAGISKELNELKGVIELLSTNYSST-----LEYYKNSSSES 1422
QY 542 TRAGYLENGKLL-----QRNIGLEIKDYOILQO--SEKEYIRIDAKVYPSKID----- 589
Db 1423 VRFQQLANGETFKAEKKNASARLAEEKLEQIVLDQSDIDDKV--KKEIGIKRE 1479
QY 590 -TKIOEALNINOEMKALGLPKYTKLTENVHN-----RASNIVESAYIILNEMK 640
Db 1480 ILKKESALTF--WEES--EKFQOMCSHMEKAKKKEKIEYLKNNGDG-----K 1526
QY 641 NNISDILKKTNYLVNDGNRFVFTDITL-----PRIAQYTHODEITYOVHSGLYV 693
Db 1527 ANITDSQMEEGVYSAEHAHFHVEAOVDTKAFCEISIVAYTKMDLNFMSLMKEVKY 1586
QY 694 PESRSILLHGSPKGVLELRNDESEGFIERGHAVDYAGYL-----LD 734
Db 1587 K-----CEKKNDE--AEKYSALKPYDGRIKARVSENEKIKSELKEKAYE 1630
QY 735 KNOG---DLVTNSKKFIDIFKEE-GSNLTSYGRT--NEAEFFAEAFRLM 777
Db 1631 KRESSQINDVSTKSLQIDNCROQLDSVLSNIGRVKQNALQFDSADKSM 1680
RESULT 4
SCPI_RAT STANDARD; PRT; 997 AA.
AC 003410;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SYNAPTONEMAL COMPLEX PROTEIN 1 (SCP-1 PROTEIN).
GN SYCP1 OR SCPI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=93099884; PubMed=1464329;
```

```
RA Meuwissen R.L.J., Offenberg H.H., Dietrich A.J., Riesewijk A.,
RA Iersel M., Heyling C.;
RT "A coiled-coil related protein specific for synapsed regions of
RT meiotic prophase chromosomes."
RL EMBL J. 11:5091-5100(1992).
CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF
CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
CC CHROMOSOMES DURING MEIOTIC PROPHASE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
CC SYNAPTONEMAL COMPLEX.
CC -1- TISSUE SPECIFICITY: TESTIS.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EXCLUSIVELY IN MEIOTIC PROPHASE
CC CELLS.
CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
CC CORRECTED IN POSITION 6 TO MAXIMIZE THE SIMILARITY WITH THE
CC OTHER SPECIES SYCP1 SEQUENCES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X67805; CAA48006.1; ALT_FRAME.
DR PIR, S28061; S28061.
KW Nuclear protein; Meiosis; Cell division; Phosphorylation;
KW DNA-binding; Coiled coil.
FT DOMAIN 108 819 COILED COIL (POTENTIAL).
FT DOMAIN 118 121 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 701 704 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT FT 902 905 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 982 990 ARG/LYS-RICH (BASIC).
SQ SEQUENCE 997 AA; 116511 MW; 229D9823FD684BE CRC64;
Query Match 5.88; Score 229.5; DB 1; Length 997;
Best Local Similarity 21.88; Pred. NO. 0.024;
Matches 192; Conservative 154; Mismatches 331; Indels 205; Gaps 48;
QY 44 VKEKKNDENKRRKDEENKTOEHLKEMKHYKIE-----VKGEAVK----- 88
Db 128 LKQENKLOENRKILTEOKRAIQELOFENKVSILKEELIOENKDLKENNATRHWCNL 187
QY 89 KEAAEKLEKVPDVLWEYKAIIGKIYI-VGDITTKH-----SLVALSEDK-----KIK 138
Db 188 KETCARSAEK--TSKYEREETRGQVYVDLNNNEKMLAEELRQVQENARLMEHRKLK 245
QY 139 DIYKQALLHEHT--YAKGGEYPLVYQSSDEYVENTEALVYVEIGKILSRDLISK 195
Db 246 EDHEKIOHLEEYQKEVKNENOVSLLIQSTEK--ENKMKDLTFLE-----ESRD---K 296
QY 196 INQYOKF-----LDVLTINKNASDSQDGLFTNQ-----LKEH---PTDFSVEF 238
Db 297 ANQLEEKTKLODENMLKELNEKKDLUTSELEDIKMSQMSMTOKTLEEDLOIAKTTIYL 356
QY 239 LEQNSNEVOEV-----FAKAFAYIE-----POHRVDVQLVAPE----- 272
Db 357 TEKEAQMEELNKAKTHTSLVLTTELKATCTLEELRTPEQRLNNEDEQKSLIMLEOKK 416
QY 273 --AFNYMDK--NEQELNLSLELK-----DQRMLSREKMEKIKQHYQHMSDLS---EE 321
Db 417 SSELEEMTKFKNNKEV--ELEELKTIILAEDQKLDEKKQVEKLAELQKQDEQLTFLQOT 474
QY 322 GRLKKLKIQIIEPKKDDIHSLSQ-----EKELLRKIQID-SSDFLSTEEKEF--- 370
```

```

Db 475 REKEIDHLEQVAVTQTSSEHYLKQVEEMKTELEKLNKIELTANSMDLLENKKLVQVE 534
Qy 371 -----LKKLOIDIDISLSEEEK-----ELLRIQVDSNPISSEKEKELK-----KLK 413
Db 535 ASDMVELLKKHODIINCKKQOEERMLKQIETLEKEKNLDELESVKEFIIOGDEYCK 594
Qy 414 LDIO-----PYDINORLODTGGLDPSINLDVRKQYKRDIONDALLHQ-----SIG 461
Db 595 LQDSSEANASIEYVLK-EKQKILLENKCNL--KKQLENSKNIEE-LHQENKALKKK 650
Qy 462 STLYNK-IYLYENMNINLTATGADLVSTDTKIN-RGIFNEKKNFXY-SISSNYMI 518
Db 651 SSANENQOLNAYE-IKYNKLELELA-----STKQKEEMINNYQKEIKIKISEKKL 701
Qy 519 VDIERNPALDNEFLKWRQISLPTTRAGY-----LENGKLLORNIIGLEIKQVOIIO 570
Db 702 GEVEKAKAVYDEAVK--LQKEIDLRQCHKIAEVVALMEKHKHQYDKIVEERDSEGLYKN 759
Qy 571 SEKEYIRIDAKVYPRKSIDPKIOE-----AQLNINQEMNALGL-PKYTLITFENHNR 624
Db 760 REEE--QSSAKVALETLSIRNELVSLKKQLEVEKEEKLMEQENTALTQDKDKKI 817
Qy 625 ASNIVESAYLLLENMK---NNIOSDLIKVTNYLVGNG-----RFVFTDITLPNIAE 674
Db 818 QASLESPE--ATSMKFDSKTPSQNISRSLSSMDQSKDNDRSLAKSAKSLSTVTK 875
Qy 675 QYTHODEIYEQVSKGLYVESRSLHGPSK-----GVELKNDSE-----GFTHFG 722
Db 876 EYI---VKTPTKSIYQENRYLPTGSGSKRKVFEDVNSSETTDLISLVE-- 928
Qy 723 HAVDDYAGYLLDKNSD--LVTSKKKFIDIFKEGSLNLTYSYG 762
Db 929 ---EDISNRINYNNPTDSHLVTPKOTPLSLSTPASFTKFG 967

RESULT 5
SCPI.MOUSE STANDARD: PRT: 993 AA.
ID SCPI.MOUSE
AC 062209: 062329: 009205: P70192:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SYNAPTONEMAL COMPLEX PROTEIN 1 (SCP-1 PROTEIN).
GN SCPI OR SCPI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1)
RP SEQUENCE FROM N.A.
RC STRAIN=CBA; TISSUE=Testis;
RX MEDLINE=96004899; PubMed=7548215;
RA Sage J., Martin L., Guzin F., Rassoulzadegan M.;
RT "cDNA sequence of the murine synaptonemal complex protein 1 (SCPI).";
RL Biochim. Biophys. Acta 1263:258-260(1995).
RN 12)
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS; TISSUE=Testis;
RA Kerr S.M., Taggart M.H., Lee M., Cooke H.J.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RN 13)
RP SEQUENCE OF 1-149 FROM N.A.
RC STRAIN=C57BL/6;
RA Sage J., Li Y., Martin L., Mattei M.-G., Guenet J.-L., Liu J.G.,
RL Hoog C., Guzin F., Rassoulzadegan M.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN 14)
RP SEQUENCE OF 95-787 FROM N.A.
RC STRAIN=ICR; TISSUE=Testis;
RA Tsuchida J., Nishina Y., Nozaki M., Uchida K., Nishimune Y.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF

```

```

CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
CC CHROMOSOMES DURING MEIOTIC PROPHASE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
CC SYNAPTONEMAL COMPLEX (BY SIMILARITY).
CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z38118; CAA86262.1; -.
DR EMBL: LA1069; AAG4514.1; ALT_INIT.
DR EMBL: U62864; AAC53335.1; -.
DR EMBL: U62860; AAC53335.1; JOINED.
DR EMBL: U62861; AAC53335.1; JOINED.
DR EMBL: U62862; AAC53335.1; JOINED.
DR EMBL: U62863; AAC53335.1; JOINED.
DR EMBL: D88539; BAA13639.1; -.
DR MGI: 105931; Sycp1.
KW Nuclear protein; Meiosis; Cell division; Phosphorylation;
KW DNA-binding; Coiled coil.
FT DOMAIN 12 97 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 104 815 COILED COIL (POTENTIAL).
FT DOMAIN 114 117 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 697 700 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 978 901 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 978 986 ARG/LYS-RICH (BASIC).
FT CONFLICT 527 527 F -> L (IN REF. 2).
SQ SEQUENCE 993 AA; 115962 MW; 1AAFN90D64FAFE6 CRC64;

Query Match 5.7%; Score 228.5; DB 1; Length 993;
Best Local Similarity 22.0%; Pred. No. 0.026;
Matches 202; Conservative 144; Mismatches 336; Indels 237; Gaps 49;

Qy 44 VKEKKKNDKRRKDEBRNKTQEBHEKIMKHVYKVEBAVK----- 88
Db 124 LKQKENKLOENRKIKIEQKRA---IQELOFENKYSLKLEEEIQENKDKLEKNNATIH 179
Qy 89 -----KEAAEKLEKVPDVLNEMKAIQGIYI--VDGDTKHI-----SLALSDEKK----- 135
Db 180 CNLKEKTCARSAEK--TNKYEYERETROYVDINSIEKMTLAFEEELRYQAEKARLEMH 237
Qy 136 -KIKDIYKDALHNEY---VYAKEGEYEPVLVIOSSDEYVENTEKALNYYVEIGKILSRD 191
Db 238 FKLEKDEKIQHLEEVQKQEVNKNKENGVSLLQISAK--ENMKKDLTFLE-----ESRD 291
Qy 192 ILKKNQYQKF-----LVLTNTIKNASDSGDGL-----LFTQKLEHPIDSEVF 238
Db 292 ---KANOLEKTKLQDENLKELESEKDKHLTSELEDIMSMQSRMSTOKALEEDLOLATKT 348
Qy 239 LEQNSNEVOE-----VFAKAFAYYE-----POHRVLOLYAP 271
Db 349 ISO-LTEVKEKQOMELKAKATTISFVYVTELKATCTILEELRTQOAPLEKNEQDLKITV 407
Qy 272 E-----AFNYMDKF--NEQINLSLEELK-----DQRMLSRYEKWEKIKQHYHMSDLS 319
Db 408 ELQKKSNELEKMTKFNKKEV--ELEELKNILAEQDKLDEKKQVEKLAELQEKBEELT 465
Qy 320 ---EEGRGLIKKIQIPIEPKKDIIHSLQE-----EKELLKRIQIDSS--DFLSTEEK 368
Db 466 FLLETREREVHDLOEQVAVTQTSSEHYLKQVEEMKTELEKLNKIELTASCMDLLENK 525

```

```

QY 369 EF-----LKKLOIDIDBSLSEKELLNRIQVSSNPLSKE-----KE 407
DQ 526 KFEVQASMALELKKHODIINCKKOEER-LIKQIE-----NLEEKHRLDELESVKE 579
QY 408 FLK-----KLKLDIOPYDINQ-----RLQDTGGLIDSPSINLVDKQYKRDIONIDL 456
DQ 580 FIQGDSEVKCKLDSSEENARSIECEVLKKKQMKLSEKCNL--KKQYENKSKIEE-L 636
QY 457 HQSIGSTLYNK-----IYENNNINNLVATGADLVSDTNTKINRGIFNEKKNF 509
DQ 637 HQE-NKTLKKSSAEIKOLNAYE-IKVSKELELEL-----ESTKQR---FEEMTNVQ 683
QY 510 -----YSSNNYMYDINERPALDNERLKWRIQLSPDTRAG-----LENGLLIQR 556
DQ 684 KEIENKKITSEKLLGVEVKAQATYDAVK--LQKEIDRCOKHIAEWALMEKKHHOYDK 741
QY 557 NIGLEIKDVOIIOSEKEEYIRIDAKVPSKIDTKIQE-----AQLNINOENKALGPK 611
DQ 742 IVEERDSLEGLYKNRQE--QSSAKIALETLSNIRNELVSLKQLEIEKEKEKLMKAK 799
QY 612 -YTKLITFNHNRVANSNIVESAYLILNEMK--NNIQSDLIKVTNYLVG-----NG 660
DQ 800 ENTALILKKKKKIQASLESPE--ATSMKFDKSTPPQONISRLSSMDGSKDNRDL 857
QY 661 RFVFTDILPNIABQYTHODEIYEQVHSK-GIYVPSRSILHGSKEVELRDNSE----- 715
DQ 858 SASAKSIPTVTYKTYKTPKSTIYORENKYIPTGSSNKKRTAFEDVNSDSSETAD 917
QY 716 --GFIFHGHAUVDDYAGYLDKQNSD--LVYNSK-----KFIDIFKEGSGN 757
DQ 918 LLSLVSE-----EDVSNLNYDNPNPDSHLVTKPQTPLSLSTPASFKFQSLKMKRDR 972
QY 758 LITSGRTNEAFEPFARL 776
DQ 973 WTIKAKIDKRRRLKEAKL 991

RESULT 6
HMMR_MOUSE STANDARD; PRT; 794 AA.
ID HMMR_MOUSE
AC 000547;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYALURONAN MEDIATED MOTILITY RECEPTOR (INTRACELLULAR HYALURONIC ACID BINDING PROTEIN) (RECEPTOR FOR HYALURONAN-MEDIATED MOTILITY).
GN HMMR OR IHABP OR RHAMM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murineae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
RC TISSUE=Lung;
RX MEDLINE=98264863; PubMed=9601097;
RA Hofmann M., Fieber C., Assmann V., Goettlicher M., Sleeman J., Plug R., Howells N., von Stein O., Ponta H., Herrlich P.;
RT "Identification of IHABP, a 95 kDa intracellular hyaluronate binding protein."
RT J. Cell Sci. 111:1673-1684(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Zhao Y., Zhang S., Turley E.;
RL submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-183 FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=129/SV;
RX MEDLINE=99107769; PubMed=9889313;
RA Fieber C., Plug R., Sleeman J., Dall P., Ponta H., Hofmann M.;
RT "Characterization of the murine gene encoding the intracellular hyaluronan receptor IHABP."
RL Gene 226:41-50(1999).
RN [4]

```

```

RP SEQUENCE OF 164-794 FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=BALB/C; TISSUE=Fibroblast;
RX MEDLINE=96011639; PubMed=7590272;
RA Entwistle J., Zhang S., Yang B., Wong C., Li Q., Hall C.L., A.J.,
RA Morat M., Greenberg A.H., Turley E.A.;
RT "Characterization of the murine gene encoding the hyaluronan receptor RHAMM."
RT Gene 163:233-238(1995).
RN [5]
RP SEQUENCE OF 318-794 FROM N.A., AND CHARACTERIZATION.
RC STRAIN=BALB/C;
RX MEDLINE=92299690; PubMed=1376732;
RA Hardwick C., Hoare K., Owens R., Holm H.P., Hook M., Moore D.,
RA Crisps V., Austen L., Nance D.W., Turley E.A.;
RT "Molecular cloning of a novel hyaluronan receptor that mediates tumor cell motility."
RT J. Cell Biol. 117:1343-1350(1992).
RN [6]
RP CHARACTERIZATION.
RX MEDLINE=94308286; PubMed=7518470;
RA Hall C.L., Wang C., Lange L.A., Turley E.A.;
RT "Hyaluronan and the hyaluronan receptor RHAMM promote focal adhesion turnover and transient tyrosine kinase activity."
RT J. Cell Biol. 126:575-586(1994).
RN [7]
RP ERK REGULATION, AND SUBCELLULAR LOCATION.
RX MEDLINE=98225222; PubMed=9556628;
RA Zhang S., Chang M.C., Zylka D., Turley S., Harrison R., Turley E.A.;
RT "The hyaluronan receptor RHAMM regulates extracellular-regulated kinase."
RT J. Biol. Chem. 273:11342-11348(1998).
RN [8]
RP REVIEW.
RX MEDLINE=99059494; PubMed=9845361;
RA Hofmann M., Assmann V., Fieber C., Sleeman J.P., Moll J., Ponta H., Hart I.R., Herrlich P.;
RT "Problems with RHAMM: a new link between surface adhesion and oncogenesis?";
RT Cell 95:591-592(1998).
RN [9]
RP FUNCTION: INVOLVED IN CELL MOTILITY. WHEN HYALURONAN BINDS TO HMMR, THE PHOSPHORYLATION OF A NUMBER OF PROTEINS, INCLUDING THE FOCAL ADHESION KINASE OCCURS. MAY ALSO BE INVOLVED IN CELLULAR TRANSFORMATION AND METASTASIS FORMATION, AND IN REGULATING EXTRACELLULAR-REGULATED KINASE (ERK) ACTIVITY.
CC -1- SUBUNIT: SUBUNIT OF THE HARC COMPLEX.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE AND CYTOPLASMIC.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; RHAMM1V4 (SHOWN HERE) AND RHAMM1; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF031932; AAC12655.1; -
DR EMBL; AF079222; AAD08670.1; -
DR EMBL; AJ005919; CAA06768.1; -
DR EMBL; AJ005920; CAA06768.1; JOINED.
DR EMBL; AJ005921; CAA06768.1; JOINED.
DR EMBL; AJ005922; CAA06768.1; JOINED.
DR EMBL; AJ005923; CAA06768.1; JOINED.
DR EMBL; AJ005924; CAA06768.1; JOINED.
DR EMBL; X64550; CAA45849.1; -
DR EMBL; X64550; CAA45848.1; -
DR MGD; MGI:104667; Hmnr.
KW Hyaluronic acid; Alternative splicing; Repeat; Glycoprotein.
FT DOMAIN 442 546 5 X 21 AA NEAR PERFECT TANDEM REPEATS.
FT REPEAT 442 462 1.
FT REPEAT 463 483 2.

```



```

FT REPEAT 484 504 3.
FT REPEAT 505 525 4.
FT REPEAT 526 546 5.
FT DOMAIN 719 729
FT CARBOHYD 741 750
FT CARBOHYD 53 53
FT CARBOHYD 134 134
FT CARBOHYD 279 279
FT CARBOHYD 446 446
FT CARBOHYD 467 467
FT CARBOHYD 488 488
FT CARBOHYD 509 509
FT CARBOHYD 530 530
FT CARBOHYD 561 601
FT CARBOHYD 601 601
FT VARSPPLIC 218 242
FT CONFLICT 19 19
FT CONFLICT 55 55
FT CONFLICT 71 71
FT CONFLICT 89 91
FT CONFLICT 94 94
FT CONFLICT 540 540
FT CONFLICT 668 668
SQ SEQUENCE 794 AA: 91799 MW: 74DB3D236224499C CRC64:

```

Query Match 5.6%; Score 223; DB 1; Length 794;

Best Local Similarity 20.1%; Pred. No. 0.032; Matches 162; Conservative 155; Mismatches 255; Indels 234; Gaps 40;

```

OY 45 KEKKNKNDKRRKDE-----RNKTOEHLKMKHIVKIEVKGGEAVKKE-----AAEK 94
DB 76 KSGQKNDRKVKLEKRIKRLALBGRQDKRIODMESELEKTEKLAANAAREKISLSASNA 135
OY 95 LLEKVPDYLEMYKAIGIYIVDGIKHSLEALSSEDKKIKIDYGDALLHEHYVA 154
DB 136 SLEKRTLETFRANELLKAK-FSEDC---HOKNMRALESLKMLKRN---KRETKRSMYAK 188
OY 155 KSGYEPVLVIOSS-E-DYVENTEKALNVYIEGKIISRLSKINQ--PQKFLDVANTIK 211
DB 189 QSGME--LKLQAMQKDLTSEKSKIYOLE--GKLVSE-KEKIDCECEKLELEYIOETS 242
OY 212 NASDSKQDLTFLNOLKEHPTDFSVFE-----LEON---SNEQVFAKFAFYIEPOH 262
DB 243 CASDQVEKCKVDAQLEEDLKEKREILSLKQSLSENIFFSQIEDLYVKC----- 293
OY 263 RYVOLIYAPEARNYMDKNEOEINISLE-ELKDQRLMSREKMEKIKOHYQHSWSLSSE 321
DB 294 ---QLETERDNLVSKDERERAEETLSAEQOILTERLALEROEYEKLOQ----- 337
OY 322 GGLGLKLIPIEPKKDDIHSLSOEKELKRIODSSDF---LSTEKEFLKLIQIDI 378
DB 338 -----KELQ-----SQSLQOEKELSAHLQOOLCSFOEEMSEKNVFEELKLAL 382
OY 379 RD-----SLSEEEK---ELNLRIQVSSNPSEKEKEFLKLIQIDOPY 419
DB 383 AELDVAQKEQESERLVQLEERKSTAEQLRTL---DNLREKEVELEKHAIAAQAQI 438
OY 420 -----DINRLQDTGLDPSINLDYRKQYKRDIONIDALLHOSIGSTLYNKIYLY 471
DB 439 LIAOEKYNPTAOSLRVYTAQLES-----VOEKYNDPTAOSL---RDVYTAQLESQEKY 487
OY 472 EN--MNINNLATLGADLVSDTNKINGINFNEFKKNKRYSSNY-----MIVLIN-E 523
DB 488 NPTAOSLRVYTAQLESQEKYNDPTAOSLRYTAQLE-----SVQEKYNDPTAOSLRVYSAQ 542
OY 524 RPALDNERLKMRIQSLPDRAGYLENGKLIQINIGLEIKDVOIIOSEKREYIRIDAKVY 583
DB 543 LESYKSTLKEIEDLK-----LEN--LTLOEKVAMAKSVEDVQO----- 580
OY 584 PASKIDTKIOEQALNINDEMNKALGLPKYTKLITFVNHNRYSNIVESAYVLLINENKNI 643
DB 581 -----QILTAE-STNOE-----YARWQ-DIQNR-----STL 605

```

```

OY 644 QSDLIKKVTNYLVDCNGRVFTDITLPNTAEQYTHODETY-EGVHSGLYVESRSIL-- 700
DB 606 KEKEIKETITSSFE-----KITDLKNQRODEDFRQLEKGRKTAKEVNMKE 655
OY 701 -----LHGSPKGYELRND-----EGFTHERGHAHV-----DDYAGYLLD 734
DB 656 LIMEIKKRLLYELYEKTRKPPQOQDAFEAKQALLNENGAIQEOQLNKIRDSYAGLLGH 715
OY 735 KNSQDVLVNSKRFIDFKREGSNLJS 760
DB 716 QNLKOKI-----KHVYKLKDNSQLKS 737

RESULT 7
ID CENE_HUMAN STANDARD; PRT; 2663 AA.
AC 002224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CENTROMERIC PROTEIN E (CENP-E PROTEIN).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE. AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC
DR EMBL: Z15005; CAA78727.1; -.
DR PIR: S28261; S28261.
DR HSSP: P03069; 121T.
DR MIM: 117143; -.
DR InterPro: IPR001752; kinesin.
DR Pfam: PF00225; kinesin; 1.
DR PRINTS: PR00380; KINESINHEAVY.

```


QY	335	PKKDDITHSLQSEKELLRKRIQIDSSDF-----LSTEE---KEFLAKQIQIDRSLSE	385
Db	543	PKRIDKY-----AKIYVDRIEDTEFEKRVANVAVNRLGVCVRFKALKDYKLEKFGDE	596
QY	386	EKELNLRIOVDSSAPLSEKER-----EFLAKTKLIDQPYD---INQRQD-----	427
Db	597	AEKNAEYVMYTHHN---EEKEIIEYKKLAKKERNSNDPNEINQIITREEFLNSENPKILI	653
QY	428	-----TGLIDSPSIN---LDVAKQYKRDQIONID-----ALHOSIG--STLY	465
Db	654	VTDMILNG--FDAPRLKVMYLDKRPYLGHRLLQALARTNRPYPDKERGLVDSYGLGKVLVLT	711
QY	466	NKIYLENNMINNLATLGLADYDSDIDNTRKINGRINERK-----KNFKYSISSN	515
Db	712	ETMALYNNMLAEEEREDFEKNNLLSIDE-----IFOEFLKLEMYKESLSKNLK--INDE	763
QY	516	YMIYDINERPAL-----DNERLKWRIQSPDRAGYLENGK-----LIIQ	555
Db	764	DLSIDVNTLTKTKKPKDFNNNELKEK---DLIAFYAEDGKNARILKLIDLKAVIKLY	819
QY	556	RNIG-----LEIKDYOIIKQSEKEYIRIDAKVVPKSIDTKIQEA-----QL	597
Db	820	KALGSYPOKIFYIEDIDLLSFI---YAVLLKTKLPPKKSNRKWEELISFINNMKMLVDDL	876
QY	598	NINQENKALGLPKRYTYLTFIPVNHNRASVIVESAYLILNEMKNNT---QSDLIKRVNY	654
Db	877	TVIEEIN--LNPDDLDLKEIKENIGKREIKRAVANYYFIL---KNSILDKQHPPIYK---	927
QY	655	LVQNGRFVFTDILPNAIE---QYTHODELYEQVHSGKGLVPSRSILHSGSGVELN	712
Db	928	-----ELERLELRKRWIMKKRIDK-----IYNALIKNLMELKN	962
QY	713	D-----SEGIHFEGHAVDVYAGYLLDKKNQSDL---VTNSKKFIDIFKEGCSNLTYSGR	763
Db	963	NYDKKIKGSSIERIKRSIYIGENILKQD-DIKLNLENTEKLI---TKQNLNKLKSK	1011
QY	764	TNEAEFFAE 772	
Db	1018	LQRRKFKKE 1026	
RESULT	9		
MYSD_DICDI		STANDARD;	PRT; 2245 AA.
AC	P54697;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	MYOSIN 1D HEAVY CHAIN.		
GN	MYOJ.		
OS	Dictyostelium discoideum (Slime mold).		
OC	Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.		
OX	NCBI_TaxID=44689;		
XM	[1]		
XM	SEQUENCE FROM N.A.		
RP	STRAIN=AX3;		
RC	MEDLINE=96215148; PubMed=8636147;		
RA	Hammer J.A. III, Jung G.		
RT	"The sequence of the dictyostelium myo J heavy chain gene predicts a		
RT	novel, dimeric, unconventional myosin with a heavy chain molecular		
RT	mass of 258 kDa."		
RL	J. Biol. Chem. 271:7120-7127(1996).		
RN	[2]		
RP	SEQUENCE OF 1-1021 FROM N.A.		
RX	MEDLINE=97039016; PubMed=8884597;		
RX	Peterson M.D., Uristoste A.S., Titus M.A.;		
RT	"Dictyostelium discoideum myoJ: a member of a broadly defined myosin		
RT	V class or a class XI unconventional myosin?";		
RL	J. Muscle Res. Cell Motil. 17:411-424(1996).		
RN	[3]		
RP	SEQUENCE OF 182-298 FROM N.A.		
RX	MEDLINE=95023928; PubMed=7937787;		
RA	Titus M.A., Kuspa A., Loomis W.F.;		

```

RT "Discovery of myosin genes by physical mapping in Dictyostelium.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9446-9450(1994).
CC -I- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN AND HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -I- SUBUNIT: HOMODIMER.
CC -I- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U42409; AAA85186.1; -.
DR EMBL: L35322; AAA79658.1; -.
DR HSP: P08799; 1MND.
DR DictyDB: DD01095; myoJ.
DR InterPro: IPR002710; myoJ.
DR InterPro: IPR000048; IQ.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF01843; DIL; 1.
DR Pfam: PF00612; IQ; 3.
DR Pfam: PF00063; myosin_head; 2.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1.
DR ProDom: PD003376; DIL; 1.
DR SMART: SM00015; IQ; 3.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 3.
KW Myosin; ATP-binding; Actin-binding; Calmodulin-binding; Repeat;
KW Coiled coil.
FT DOMAIN 1 809 MYOSIN HEAD-LIKE.
FT DOMAIN 824 851 IQ 1.
FT DOMAIN 872 901 IQ 2.
FT DOMAIN 943 972 IQ 3.
FT DOMAIN 973 1812 COILED COIL (POTENTIAL).
FT DOMAIN 1813 2245 TAIL.
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 669 749 ACTIN-BINDING.
FT CONFLICT 191 191 L -> F (IN REF. 2).
FT CONFLICT 284 284 A -> T (IN REF. 2).
FT CONFLICT 291 291 G -> R (IN REF. 2).
FT CONFLICT 332 347 NKGGEFEEGVSDDEH -> IHWMEELVYRMS (IN
FT REF. 2).
FT CONFLICT 550 550 N -> K (IN REF. 2).
FT CONFLICT 865 866 HH -> QQ (IN REF. 2).
SQ SEQUENCE 2245 AA; 258478 MW; 615E5EFLIDAB4SBE CRC64;

Query Match 5.4%; Score 215; DB 1; Length 2245;
Best Local Similarity 19.4%; Pred. No. 0.22;
Matches 169; Conservative 147; Mismatches 274; Indels 280; Gaps 40;

QY 50 NKDEKKRKDEER-----NKQOE---HLKELMKHYV--IEVGGE-- 85
DB 1123 MKQEPDRLSQEDRTDNTNNQLIEIOLKKNASTLEEDYFSLGIRDLEROVELERDENOL 1182
QY 86 -----AVKKEAEKLE-----KVPSDVLEMYKAIGKIIYVG 119
DB 1183 IKERDLSLGQSSQSQFSGAALERKQOLEVLQVQDSOEDLIKLSSEKLGSEEAQKQINQLEL 1242
QY 120 DITKH-----ISLEALSDEKKRKIKDYLKEDKALLHEHYVYAKGEYEPVLVIQSSDEVYNT 174
DB 1243 ELTDHKSKLQIQLOLQTEGSENEKIKTKLKG--LEEVQDEKQKQOLEERIKQSKQSVDE 1299
QY 175 EKALNVEYEIGIISRDILSKINQYQKRLVDLVNLTIKNASDSG----- 218
DB 1300 KNSLTQLTTLVVEESTQVSTNVSHQEKITTLTKSTLEELNKSIGKLQAEQKNKDEIRKI 1359

```


QY 753 EEGSNLTSYGRTEAEFFAEAFRL 776
DB 1559 EKSNAVKRLSTESKEITLQPNRSL 1582

RESULT 11
Y109_YEAST STANDARD: PRT: 1679 AA.

AC P40457;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 01-FEB-1995 (Rel. 31, Last annotation update)
DE HYPOTHETICAL 195.1 KDA PROTEIN IN DNA43-UB11 INTERGENIC REGION.
GN Y11449C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajadream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z38059; CAA6129.1; -
DR PIR: S48385; S48385.
DR SGD: S0001411; MUP2.
KW Hypothetical protein.
SQ SEQUENCE 1679 AA; 195141 MW; 298950CC52202D8F CRC64;

Query Match 5.2%; Score 209; DB 1; Length 1679;
Best Local Similarity 19.8%; Pred. No. 0.27;
Matches 171; Conservative 155; Mismatches 272; Indels 264; Gaps 42;

QY 44 VEKEKNKDEKRNKDEER---NKTQEHKLEIKMHIKIVKGEBA-----VKKEA 92
DB 644 IYKKSCKKTKTLEDPENKGLAKEKRMLEAIDHL-KAELEKQSWPSYIHVEKERA 702
QY 93 EKLEKVPDVLWKYKAIGKITYVDGITKHISLEALSDKKKIND---IYGDALL 147
DB 703 STELSQ-----SRIKIKSLEVEISKLEKKTASFIPTKESLT 738
QY 148 HEHYVAKGEYPLVYIOSSSEYVENTEKALNYYIIGILSR-----DILSKINPYQK 202
DB 739 RPEQCCCKKELQMRKLESE--ISHNENKMPFSSKEGQYKAKIKELNNLERLRSDDLQ 796
QY 203 PLVDLNTINASDS-----DGODLLFTNOLKE-----HP 221
DB 797 KIOEIESISCKDSOLKAMQNTIDTEMKMSLITLSKNTTIEKLSSEIENLDKELK 856
QY 232 TDFSEVELEFONSNEVOEYAKAFAYIEPOHR---DVIOLYAPEAFNYMDKENE--QEIIN 286
DB 857 TFFQKFLQNSD-----ASTLEPTLRKELEQIOVOLKANSQIOAVEIISNNE 906
QY 287 LLELEKQDM-----LSREKWEKIKQHYQHMSDSISESGRLKLLQIPIEP 335
DB 907 NLLIEELNLAATKENYDAKIELEKKEKWAR-----EEDLSRL-RGELGIR-ALQP 956
QY 336 K-KDIIHLSQOEK--ELLKRIO--IDSSDFLSTEEKEFLKLQIDIRDSISEEK--- 387

DB 957 KKEGALHFVQOSEKRLNERYIQKMIETKEKST-IYQCKKKEKSOYSTKKEKDS 1015
QY 388 ELLNRIQVDSNPLSEKEKEFLKLDIQPYDINORDTGLDPSFIN-LDVKKQYK 446
DB 1016 ELVIRLEKDAADQAE-----LTKTK-----SLVSAQDLDPKHEKWEKAE 1061
QY 447 RD-IQVIALHOSIGSTLYENKININNTLTGADLVDSIDNKRINKINGINEK 505
DB 1062 RELISNIEQT-----ESLRVEN-----SVLIEKVDTPAANNDDHDK 1099
QY 506 KNEFYSSNYMVIDNERPALDNERLKMRIQSPPTFRAGYLENGKLIIRNIGLE--IK 563
DB 1100 ---LVSLFSLR-----HERSLETKLTCKRELA-----FVQKNSLEKTIN 1140
QY 564 DVOIIO-SEKEYRIDAKVPSKIDTYIQEAQNLINQEWKALGPLYTKLITFNVHN 622
DB 1141 DLQRTQTLSEKEY-QCSAVIIDEFKDITK-EYQVNIKLE-NNAI-LQSKLVTEKNE 1196
QY 623 RVA-----SNIV-----ESAYVLINEMKNNIQSD 646
DB 1197 IYQLNDROEISRLQDRLIQKQVYSINSKILVYSEMEQCKORYODLSQOKDAOKK 1256
QY 647 LIKKVTNYLVQNGRFRVETDITLPNIAEQYTHODEIYGVHVK----- 689
DB 1257 DIEKLTNEISDLKGLKSSAENANADLENK---NRLKQAHEKLDASKKQOALTNELE 1313
QY 690 -----GLVPEPSRILLGPPSKGVELRNDSEGFIFHEGHAVDVAGYLLDRNQSD 739
DB 1314 LKAIKKLEQDLDFENAKVIDLDTKLKAHELQSEDSVSDHE-----KDTY-----RTLME 1363
QY 740 LVYNSKKFLDIFKEEGSNLTSY 761
DB 1364 EIESLKEQLDIFKTANSSDAF 1385

RESULT 12
Y104_YEAST STANDARD: PRT: 691 AA.

AC P40460;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL 80.5 KDA PROTEIN IN SLN1-RA25 INTERGENIC REGION.
GN Y11444W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajadream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z38059; CAA6134.1; -
DR PIR: S48390; S48390.
DR SGD: S0001406; TID3.
KW Hypothetical protein.

SEQUENCE 691 AA; 80487 MM; 7DBC492227A80093 CRC64;

Query Match
Best Local Similarity 5.2%; Score 208.5; DB 1; Length 691;
Matches 132; Conservative 117; Mismatches 195; Indels 201; Gaps 32;

QY 119 GDTKHSLEALSDKKIKIYKGDALLHEHYAAK-----EGYEPV-----161
D 105 GSLQHLST-----NRDPRLRDKNFQSAIQEETIYDLKKNKFDIEIHHISIKFLKOPTQK 160
QY 162 -----LVIOSSDYENTKALNVYEIGKILSRILEKINP-----YQKFL 204
D 161 GFTIIFKVLRLDPGV-GFTKSIENETIQLKRLRYFELSLINSQISAVGSGNMHFL 219
QY 205 DVL-----NT-----IKNASDSDGQ-----DLFTNQ 227
D 220 GMLHMVATNIKLMCLNKVDRSLINQTOETILSOPKLTLDQDQOQERVELMVKRL 279
QY 228 KEHPTDSVER--LEQNSNEVOEYFAKAFAYIEPHRDVQLYAPAFNYMDKFNBOEI 285
D 280 IDYTESYKSLKLEEDNEPSEMOLKLEFEKVFHINTDINLOT-QNDNLYERY--QEV 336
QY 286 NLSEELKQDMLRYEKWEKIK--OHYHMSDL--SEEGRLKLLQIPTEPKKD 339
D 337 MKISOKITTR-----EKMKALKSDSKNYENYVNAAMKQSQEMFCKLEKMSCECELEE 391
QY 340 ITHLSOEKELKRIQIDSDFLSTEKEFLKLLQIDIRLSSEKELLRIOVDSN 399
D 392 -IKALQSNISLHKILR---KKGISTEOLF-----LQNOERELTR-ELDKIN 434
QY 400 PLSEKEKELKLLDQPYDINQRLDTGGLIDSPSINLDVRYKQKIDONI-----452
D 435 IQSKLTSSISKRLKEA-----GIEFS--LDLTKQYSSIONLRRSRQL 479
QY 453 -----DALHOSIGSTLYNKIYLENNINMLTATGADLVSDTNT--KINRGIFNEFK 505
D 480 GHNVNDSLKKNISENLDRDF-HGDISYBQLPKR-GSGINSEIKSLIKLDELQOEIK 537
QY 506 -----KNFYSSISNMYIVDINERPALDNERLKWRIQLSPDRAGYLENGKL 552
D 538 TIERKDNITLEKDIKNKLKHDINEK--TQINKELELESEANSKPELSQK-----ENRRL 588
QY 553 ILQNGIGLEIDVOIIOSEKEYIRIDAKVYPRKSI-----DKRIOEALQINQOE 602
D 589 LVAORIELEKKEKKI-----NDSNLLMKTKRISDAEELVSTELKLELVADLNRK 638
QY 603 WNKALGLPKYTKLITFVNHRYNRYASVAYLLINEMKNINQSD 647
D 639 -----RYKLLHQQ-VIHHVID-----ITSKFKINQSSSL 664

RESULT 13
SMC3_YEAST
ID SMC3_YEAST STANDARD; PRT; 1230 AA.
AC P47037;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CHROMOSOME SEGREGATION PROTEIN SMC3 (DA-BOX PROTEIN SMC3).
OS SMC3 OR YJL074C OR J1049.
GN Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M303;
RC MEDLINE=97474309; Pubmed=9335333;
RA Michaelis C., Ciosk R., Nasmyth K.;
RT "Cohesins: chromosomal proteins that prevent premature separation of
RT sister chromatids";
RL Cell 91:35-45(1997).

RN [2]
RP SEQUENCE FROM N.A.
RA Rose M., Koeltter P., Entian K.D.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Sor F.J.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Y14278; CA474655.1; -;
DR EMBL: Z49349; CA89366.1; -;
DR EMBL: X88851; CA61313.1; -;
DR SGD: S0003610; SMC3.
DR InterPro: IPR003439; ABC_Transport.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02483; SMC_C; 1.
DR Pfam: PF02463; SMC_N; 1.
KW Mitosis; ATP-binding; Coiled coil; Nuclear protein.
FT NP_BIND 32 39 ATP (POTENTIAL).
FT NP_BIND 172 482 COILED COIL (POTENTIAL).
FT DOMAIN 685 1041 COILED COIL (POTENTIAL).
FT DOMAIN
SQ SEQUENCE 1230 AA; 141336 MM; B152D86F7780341F CRC64;

Query Match
Best Local Similarity 5.2%; Score 208; DB 1; Length 1230;
Matches 162; Conservative 146; Mismatches 256; Indels 242; Gaps 39;

QY 57 KDEERNKTOEEHLKEIMKHVIEKGEA--VKKAELKLEKVPDVLKEMYKIGKI 114
D 158 KDKERLQLE-----DVYGAKSFEVKIKASIKKMEETQKKIQINKMG---201
QY 115 YIVGDIKHSLEALSDKKIK--DIYGDALLHEHYAAKGYEPV-----161
D 202 -----ELNSKLSMDQERKELKYNELERNKRYOFTLYDRELNEVINOMERLDGYN 254
QY 162 LVIOSSDYENTKALNVYEIGKILSRDILSKINOPYKFLDVLNTIKNASDSDGDL 221
D 255 NTVYSSEGIQELDKREDMDQVSKLSS-----EASLKINKATDLO-QAK 300
QY 222 LFTNQLKEHPTDSVEYFLEQNSNEVOEYFAKAFAYIEPHRDVQLYAPAFNYMDKFN 281
D 301 LRESEISQKLTNVWKI-----KDVQO-----QIESNEEQNRLDSAT 337
QY 282 EOEINLSLEELKDO--RMLSEKWEKIKOHYHMSDSLSEEGRLKLL--LQIPLEP 335
D 338 LKEIKSITTEOKKSLKLPRIYQELTYEEANMYKQLASLQOKDILIKGEYARFKSKD 397
QY 336 KKDDIHSLSQEEK-----ELKRIQIDSDFLSTEKEFLKLLQIDIRLSLSEEEKE 388
D 398 EROTWIHSEIEELKSSIQNNLESLQIQMDRTSL--RKQY-----SAIDEELEE 444
QY 389 LLNRIO-VDSGNPLSEKEKELKLLDQPYDINQRLDTGGLIDSPSINLDVRYKQYKR 447
D 445 LIDISNGPDTKGLEDDPSLDI-----HLKQKLSE-----SLDTRKELMR 484
QY 448 DIQNDALHOSIGSTLYNKIYLENNINMLTATGADLVDS-----TDNFKINR--498

```
Db 485 KEOKL-----OTVLETLTLDV-----NONORNVNETMSRLANGIINVKEITEKLSIPES 535
Qy 499 --GTFNE-FKKNFKS-----ISSN---YMYDINERPAL-DNE-RLK-WAI-----Q 537
Db 536 VFGTIGELIKYNDKRYKCAEYIGNSLPHIVDTEATLLINNELYRKGKGVTEPIPLNR 595
Qy 538 LSPDTRAGYLEN-----GKLLIQRNIGLEIKDVQIJK 569
Db 596 LSLSDVAFPSNTTQIOFTPLIKIKYEPREKAVKHVFGTTIVYKDIGGLK---LAK 652
Qy 570 QSEKEYIRIDAKVVPKSIDTKIOEAOLINQENKALGLPYTKLITFNHNRYS--- 626
Db 653 KHKLNATILDDGRADKRGVLT-----GGYLDQHKRTLESJLKNLNESSQHK 699
Qy 627 NIVESAVYILNEMKNINOSDLIKKYTNLVDSNGRFVFTD--ITLPNTAEQYTHODEIYE 684
Db 700 KILEELDFVRND-----LNDIDTKIDQ--VNGNIRKVSNDRESVLNTI----- 740
Qy 685 OVHSKGLVVPESRSIILHGPSKGYELR---NDSEGFIEFHADVADYAGVLLDKNQSDL 740
Db 741 EYRTSLTKKNEKILTEESLNATILKLEKLTNTNTFAOE---KLNTENDLLOEFDSEL 797
Qy 741 VTNSKRFIDIFEKESG-----NLTS 760
Db 798 SKEKERLESJLTKESIAHNKLNITS 823

RESULT 14
Y328_MYCGE STANDARD: PRT: 756 AA.
AC 049419: 049308: 049320:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MG328.
GN MG328.
OS Mycoplasma genitalium.
OC Bacteria: Firmicutes: Bacillus/Clostridium group; Mollicutes.
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 33530 / G-37;
MEDLINE=96026346; PubMed=756993;
RA Fritscher C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fritschmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Feldman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Ueberback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier J.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL "Natl. Acad. Sci. USA 92:10760-10764 (1995).";
RN [2]
RP SEQUENCE OF 19-113 AND 155-272 FROM N.A.
RX STRAIN-ATCC 33530 / G-37;
MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U39714; AAC71552.1;
CC EMBL; U02203; AAD12492.1;
CC EMBL; U02188; AAD12472.1;
CC DR
```

```
DR TIGR; MG328;
KM Hypothetical protein: Complete proteome.
FT CONFLICT 52 52 T->R (IN REF. 2).
FT CONFLICT 65 65 S->W (IN REF. 2).
SQ SEQUENCE 756 AA; 88407 MW; 1D3AD4EFOAB5C068 CRC64;

Query Match 5.2%; Score 205.5; DB 1; Length 756;
Best Local Similarity 22.2%; Pred. No. 0.15; Indels 219; Gaps 42;
Matches 158; Conservative 120; Mismatches 215;

Qy 46 EKEKN-----KDEKRRDE-----ERNTOEHLKEIKHIVKIEKGEAV 87
Db 61 EKEKSTVSFENMIDLVKDKKDKVNLHNYANOLKTRREBOGYVQTMAFOEVDOSLV 120
Qy 88 KKEAAEKLEKVP-----DYLEMYKAIGKIYVYDGTITHSLEAL-----SED 133
Db 121 TKEITGFQELKPSVYTAEDKDE-----IKPATKQVSPFEELFNQPSSEINET 169
Qy 134 KKKIDYKQKALLHENEHYAAKEGYEPVLYIOSSEY--VENTERKLNYYEIKGLSRD 191
Db 170 KRPEVQIFSTDRV-----REP---EQFDFYSIEMLTKAIN--PVHKTIOYD 211
Qy 192 ILSKINQPY--OKFLDVLNTIKNASPD---GODLLFTN-QLKEHPTDSVEFLQNSME 245
Db 212 --QNDQPFVVKRIKLEQHPTKKVDLDYNNKELLEBNADLK-----QIDDKENND 264
Qy 246 VOEVFAKAFAYYIEPQHRVLYOLYAEAFNYDKFNEQENISLEELKDQRMLSRYEKWE 305
Db 265 --QIF-----DL-----EQEI---DDLK--RRISE-----E 283
Qy 306 KIKQ--HYQWSDSLSEEGGLKILQ--IPLEKKDIDIHLS--QEEKELKRQIOS 359
Db 284 KSKHLTKKLQDLQENLDYEQLNKPVALNPISDEVNEELENKQEKALSD-QLDA 342
Qy 360 SDFLSTEEKEFLKLIQIDIRDSLEEEKELNRIQYDSSNPISSEKEFEFLK--KILDI 416
Db 343 LNKSSNVQOQLALPVP--LNNQINELQNLQTLAREANQNLQDLYEQNDPLKMKELHON 401
Qy 417 QPYDINQRLQDTGG---LIDSPINLDVRKQYKRQIDQNLIDALHOSISTLYNKKIYLYE 472
Db 402 TSNDENEKYYDDLNOYELFEDENETKFD--KIQOQALNLD--YOKTISALKHE----- 452
Qy 473 NNNINNLTTGLADVDSDNTKINQIFNEFKKKNKYSISSYMYLVQDINERPALD---- 528
Db 453 ---NDV-----LIDETEMTRSKDNDFNNTKNSP-----EOKKALDEKLN 489
Qy 529 -----NERLKWRI-QLSP--DTRAGYLENGKLLIQRNIGLEIKDVQIKOSEKEYIRIDAK 581
Db 490 GLTIQNOQLQDRIAELEBENNEKSNLNTQOLV--NLQOQLKDSQMLFN----- 535
Qy 582 VVPKSIDTKIOEAOLINQENM---KALGLPKYTKLTFFVNHNRYSNIVESAVYILN 637
Db 536 --VAQDKLAT--LEEVNLALNEKINDLEDELGS-----SENSNNILAK--LQADHEILQ 583
Qy 638 EKKNNIQSDLIKVTNYLYVDGNGRFVFTITLPNTAEQYTHODEIYEOVHSK 689
Db 584 ESYGKLKTDPERKLKKNKLNDA-----EYQVODLSAFETNSP 621

RESULT 15
MLP1_YEAST STANDARD: PRT: 1875 AA.
AC 002455:
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN-LIKE PROTEIN MLP1.
GN MLP1 OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RX MEDLINE=93247549; PubMed=8483450;
RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Gen. Genet. 237:359-369(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,620 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MPT1 genes and three
RT new open reading frames.";
RL Yeast 9:1349-1354(1993).
CC -1- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC -1- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -1- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPT1".
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L01992; AAA34783.1; -
DR EMBL: X73541; CA51948.1; -
DR EMBL: Z28320; CA82174.1; -
DR PIR: S38173; S38173.
DR SGD: S0001803; MPT1.
KM Coiled coil; DNA repair.
FT DOMAIN 69 487 COILED COIL (POTENTIAL).
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
FT CONFLICT 301 301 R -> A (IN REF. 1).
SQ SEQUENCE 1875 AA; 218455 MW; 683AD34C9066867 CRC64;

Query Match 5.28; Score 205.5; DB 1; Length 1875;
Best Local Similarity 20.44; Pred. No. 0.42; 291; Indels 269; Gaps 51;
Matches 186; Conservative 164; Mismatches 291;

QY 2 NIKKEFIKVI-----SMSCLYTAI-TLSGPYRIPLYOGAGHDVGMHKEKKNKDEKRR 56
DB 765 NLKQELNKLSPKSDSLRMVQLQTLQ-----KEREDLLEETRRK 803
QY 57 KDEERNKTOEHLKEIMK-----HIYKIEVKG-----EAVKKEAEKLEKVV 99
DB 804 SCQKKIDELDELSELKETSQKDHHIKQLEEDNNSNIEMYNQNKIEALKKD-YESVITSV 862
QY 100 PSDYLEWYKAIGCKIYVGDY-----TKHISLEALSDEKKIKKIDYKDALLHEHY 151
DB 863 DSKQTDIEK-LQYVVKSLKEIEEDKIRLHTYVWDETINDSLR-KELEKSKINTLTDAY 920
QY 152 VYAKEGEYPLVIOSEDEYVENTEKALNVYEIGKILSRDLISKINQPYQKFLDVLNTIK 211
DB 921 SQIKE-----YKDLVFTTSQSL-----QQTNSKLD---ESFKDFTNDRK 956
QY 212 NASD-----SDGQDL-----FTNOL-----KEHPTFS--VEFLQNSNEVQEVFA 251
DB 957 NLDEKTSLEDKISLKEQMFNLNNEIDLQKKGEKADPKKRISILQNNKKEVAVKS 1016
QY 252 KAFAYITEPQH-RVYLQLYAPEA-FNYWDRKNEQ-EINLSLEELKDQ-----RML 298
DB 1017 EYESKLKIQNDLDOQITVYANTQNNYEQELQKHADVSKTISELREQLHTYKGOVKTLNL 1076
QY 299 SRYEKMEKIKOHYQHS-----DSLSEGRGLKLRQIPIEPKRD--- 338
DB 1077 SRDQLENAKLENKESWSSQKESLLEQLDLSNRSRIEDLSQNKLLYDQIQIITTAADKEVNN 1136

```

```

QY 339 -----DIHISLSQEKELLRKIQDSSDFLSTEEKEFLKLOI-DIRDSL--SEEK 387
DB 1137 STNGPGLNLTILFLRRERDLDFTVVAERDAKMLRQKISLMDVELQDARTKLDNSRYEK 1196
QY 388 E-----LNRIQVDSNPL--SEKEKEF--LKLKLDIOP--- 418
DB 1197 ENHSSIIQOHDDIMEKLNQNLRLRESNITLRLNELNNNNKKEQLQSELDRKQNVAPIES 1256
QY 419 -----YDINORLODTCGLDPSINLDVVRKQ---YKRDIONIDALLHOSIGSTLYNKI 468
DB 1257 ELTKLTKSMQEKQEE-----LKLAKYEHKRRKKSQDI-LEKHQQLSSDYEKL 1304
QY 469 YLYENNMNNLTATL-----GADLVSDTNTKINRGIFNEFKNFYYS-ISSNMYIVDI 521
DB 1305 -----ESEIENLKEELEKKEKRGAEAE-----KFNRLRRQAQERLTKSKLSQDSLTEQV 1354
QY 522 NE-----RPALDNEKLKRRIQLS-PDTRAGYLENGKLLQNRNIGLE-IKDVQIIKOSKEY 575
DB 1355 NSLRDANKVLEN-----SLSEANARIIEELQNAK-VAQGNNOLEALIRKLO--EDAEKAS 1404
QY 576 IRIDAKVVPKSIDTKIQEAOLN-----INQENKALGLPKYTKLITFENVHNRVYASN 627
DB 1405 RELOAKL-----EESTYESTINGLNEETITTLKEIEIKQKQIQOQLQATISANEQD-LSN 1459
QY 628 IVESAYVILNEMKNNIQSDLIKRVYNYLVONGRFEV-----TDITLPNTAEQY--T 677
DB 1460 IVES-----MKSPFEDKIKFIKEKTQEVNKEKILBAQERLQNPNSINNEEIKKKWESE 1512
QY 678 HQDEIYEQVH-----SKGLYVPESRSILHGPSKGYELNDSGFIHEGHNADVQAGY 731
DB 1513 HEOEVSOKIRPEALAKRRIRLPTREKI-----NKITERR--KEELEKEFEKVEERIKS 1565
QY 732 LLDKNSQDLV 741
DB 1566 MEQSGEIDV 1575

```

Search completed: December 2, 2001, 13:51:46
Job time: 341 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:46:00 : Search time 55.64 Seconds
(without alignments)
327.196 Million cell updates/sec

Title: US-09-747-521-2
Perfect score: 4145
Sequence: 1 MNKKFKIVKISMSCLVTAL.....KNAPKTFQINDQIKFIINS 809

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/6CTUS.COMB.pep:*
7: /cgn2_6/ptodata/2/1aa/6ackfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3983	96.1	776	1	US-08-021-601-2 Sequence 2, Appl1
2	3983	96.1	776	1	US-08-082-849B-2 Sequence 2, Appl1
3	3983	96.1	776	5	PCT-US94-01624-2 Sequence 2, Appl1
4	1309	31.6	472	1	US-08-021-601-8 Sequence 8, Appl1
5	1309	31.6	472	1	US-08-082-849B-8 Sequence 8, Appl1
6	1309	31.6	472	5	PCT-US94-01624-8 Sequence 8, Appl1
7	1307	31.5	508	1	US-08-021-601-10 Sequence 10, Appl1
8	1307	31.5	508	1	US-08-082-849B-10 Sequence 10, Appl1
9	1307	31.5	508	5	PCT-US94-01624-10 Sequence 10, Appl1
10	1306	31.5	456	1	US-08-021-601-6 Sequence 6, Appl1
11	1306	31.5	456	1	US-08-082-849B-6 Sequence 6, Appl1
12	1306	31.5	456	5	PCT-US94-01624-6 Sequence 6, Appl1
13	504	12.2	800	6	5183745-3 Patent No. 5183745
14	228.5	5.5	186	4	US-08-973-462-8 Sequence 8, Appl1
15	208	5.0	2710	1	US-08-480-604A-6 Sequence 6, Appl1
16	208	5.0	2710	2	US-08-405-496A-6 Sequence 6, Appl1
17	208	5.0	2710	4	US-08-915-136-6 Sequence 6, Appl1
18	195	4.7	1588	5	PCT-US83-07261-11 Sequence 11, Appl1
19	195	4.7	1588	5	PCT-US93-07261-16 Sequence 16, Appl1
20	194.5	4.7	3248	1	US-08-353-700-1 Sequence 1, Appl1
21	194.5	4.7	3248	5	PCT-US95-16216-1 Sequence 1, Appl1
22	191.5	4.6	1164	4	US-08-923-992A-2 Sequence 2, Appl1
23	189	4.6	663	4	US-09-196-293-5 Sequence 5, Appl1
24	189	4.6	663	4	US-08-209-603B-5 Sequence 5, Appl1
25	188	4.5	576	4	US-09-104-324B-4 Sequence 4, Appl1
26	187	4.5	663	4	US-08-235-836C-78 Sequence 78, Appl1
27	186.5	4.5	990	2	US-08-645-193B-15 Sequence 15, Appl1

ALIGNMENTS

28	186.5	4.5	1098	4	US-08-923-992A-8	Sequence 8, Appl1
29	186.5	4.5	1164	4	US-08-923-992A-10	Sequence 10, Appl1
30	186	4.5	2285	4	US-09-308-375-2	Sequence 2, Appl1
31	184.5	4.5	1388	2	US-08-685-576-4	Sequence 4, Appl1
32	182.5	4.4	990	2	US-08-392-625-20	Sequence 20, Appl1
33	182.5	4.4	990	2	US-08-466-961A-20	Sequence 20, Appl1
34	182	4.4	2482	1	US-08-328-254-6	Sequence 6, Appl1
35	181	4.4	663	4	US-08-235-836C-70	Sequence 70, Appl1
36	181	4.4	1312	2	US-08-592-126-148	Sequence 148, App
37	181	4.4	1312	2	US-08-687-080-51	Sequence 51, Appl1
38	180.5	4.4	1128	4	US-08-923-992A-6	Sequence 6, Appl1
39	180	4.3	693	4	US-08-235-836C-72	Sequence 72, Appl1
40	179	4.3	1104	4	US-08-923-992A-4	Sequence 4, Appl1
41	176.5	4.3	1388	2	US-08-685-576-1	Sequence 1, Appl1
42	172.5	4.2	1073	4	US-09-541-782-6	Sequence 6, Appl1
43	171	4.1	872	1	US-08-766-014-2	Sequence 2, Appl1
44	168.5	4.1	700	4	US-08-235-836C-74	Sequence 74, Appl1
45	168.5	4.1	1038	4	US-09-541-782-4	Sequence 4, Appl1

RESULT 1
US-08-021-601-2
Sequence 2, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Kilmpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Atora, Naveen
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratl, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414,057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 776 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-2
Query Match 96.18; Score 3983; DB 1; Length 776;
Best Local Similarity 100.0%; Pred. No. 66-239;
Matches 776; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 34 AGGIGDVGMVKKKKKKKKRDEERNKKTQEHKLKHKHIVKQEBANVKKKAAE 93

```

Db      1 AGGHGDVGMHVKKEKKNDEKRNKQOEHLKEIMKHIVKEVGGEAVKKEAAE 60
QY      94 KLEKVPDVLVEMYKAIGKIYIVGDTTKHISLEALSEDKKIKIDYKDALLHEHYV 153
Db      61 KLEKVPDVLVEMYKAIGKIYIVGDTTKHISLEALSEDKKIKIDYKDALLHEHYV 120
QY      154 AKEGEPVLVIOSSBDYVENTEKALNVYVEIGKILSRDILSKINOPYKFLDVLNTIKNA 213
Db      121 AKEGEPVLVIOSSBDYVENTEKALNVYVEIGKILSRDILSKINOPYKFLDVLNTIKNA 180
QY      214 SDSGQDILLFTNOLKEHPTDSVEFLQONSNVEQVFAKAFAYYTEPOHRYDLQYABEA 273
Db      181 SDSGQDILLFTNOLKEHPTDSVEFLQONSNVEQVFAKAFAYYTEPOHRYDLQYABEA 240
QY      274 FNYMDFNEOEINLSLEELKDQRMLSRYEKMEKIKQOHQMSDSISEGRLKKLQIPI 333
Db      241 FNYMDFNEOEINLSLEELKDQRMLSRYEKMEKIKQOHQMSDSISEGRLKKLQIPI 300
QY      334 EPKDDIHSLSQEEKELKRIQIDSSDPLSTEEKEFLKKLQIDIRDSLSEEEKELNRI 393
Db      301 EPKDDIHSLSQEEKELKRIQIDSSDPLSTEEKEFLKKLQIDIRDSLSEEEKELNRI 360
QY      394 QVDSNPULSEKEKEFLKKLQIDIPYDINQRLQDTGGLIDSPSINLDVRYKQYKRDIQND 453
Db      361 QVDSNPULSEKEKEFLKKLQIDIPYDINQRLQDTGGLIDSPSINLDVRYKQYKRDIQND 420
QY      454 ALLHOSISTLYNKIYLENNMINNL/TATLGADLVSDNTNFKINGINFEKKNFKYSIS 513
Db      421 ALLHOSISTLYNKIYLENNMINNL/TATLGADLVSDNTNFKINGINFEKKNFKYSIS 480
QY      514 SNYMTVDINERPALDNERLKWRIQSPTRAGYLENGKLILORNIGLEIKDVOIIOKSEK 573
Db      481 SNYMTVDINERPALDNERLKWRIQSPTRAGYLENGKLILORNIGLEIKDVOIIOKSEK 540
QY      574 EYIRIDAVVPKSKIDTQIOEAQLINOEWNKALGLPKYTKLITFNVHNRVNASNIVESAY 633
Db      541 EYIRIDAVVPKSKIDTQIOEAQLINOEWNKALGLPKYTKLITFNVHNRVNASNIVESAY 600
QY      634 LILNEMKNKIQSDLIKKTNTNYLDGNGRFVFTDITLPNIAEQYTHODEIYEVHSGKLYV 693
Db      601 LILNEMKNKIQSDLIKKTNTNYLDGNGRFVFTDITLPNIAEQYTHODEIYEVHSGKLYV 660
QY      694 PESRSILHSGSKGYELANDSEGFTHFEGHAVDYAGYLLDKNOSDLYTNSKKFLDIKE 753
Db      661 PESRSILHSGSKGYELANDSEGFTHFEGHAVDYAGYLLDKNOSDLYTNSKKFLDIKE 720
QY      754 BGSNLTSGRTNEAFEPFAFRMLHSTHAEKLYOKNAPKPEOFINOIKFIINS 809
Db      721 BGSNLTSGRTNEAFEPFAFRMLHSTHAEKLYOKNAPKPEOFINOIKFIINS 776

RESULT 2
US-08-082-849B-2
: Sequence 2, Application US/08082849B
: Patent No. 5677274
: GENERAL INFORMATION:
: APPLICANT: Leppia, Stephen H.
: APPLICANT: Klimpel, Kurt R.
: APPLICANT: Arota, Naveen
: APPLICANT: Singh, Yogendra
: APPLICANT: Nichols, Peter J.
: TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
: NUMBER OF SEQUENCES: 35
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:

```

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/082,849B
: FILING DATE: 25-JUN-1993
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/021,601
: FILING DATE: 12-FEB-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Weber, Kenneth A.
: REGISTRATION NUMBER: 31,677
: REFERENCE/DOCKET NUMBER: 15280-161-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 776 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-082-849B-2

Query Match          96.1%; Score 3983; DB 1; Length 776;
Best Local Similarity 100.0%; Pred. No. 6e-239;
Matches 776; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      34 AGGHGDVGMHVKKEKKNDEKRNKQOEHLKEIMKHIVKEVGGEAVKKEAAE 93
Db      1 AGGHGDVGMHVKKEKKNDEKRNKQOEHLKEIMKHIVKEVGGEAVKKEAAE 60
QY      94 KLEKVPDVLVEMYKAIGKIYIVGDTTKHISLEALSEDKKIKIDYKDALLHEHYV 153
Db      61 KLEKVPDVLVEMYKAIGKIYIVGDTTKHISLEALSEDKKIKIDYKDALLHEHYV 120
QY      154 AKEGEPVLVIOSSBDYVENTEKALNVYVEIGKILSRDILSKINOPYKFLDVLNTIKNA 213
Db      121 AKEGEPVLVIOSSBDYVENTEKALNVYVEIGKILSRDILSKINOPYKFLDVLNTIKNA 180
QY      214 SDSGQDILLFTNOLKEHPTDSVEFLQONSNVEQVFAKAFAYYTEPOHRYDLQYABEA 273
Db      181 SDSGQDILLFTNOLKEHPTDSVEFLQONSNVEQVFAKAFAYYTEPOHRYDLQYABEA 240
QY      274 FNYMDFNEOEINLSLEELKDQRMLSRYEKMEKIKQOHQMSDSISEGRLKKLQIPI 333
Db      241 FNYMDFNEOEINLSLEELKDQRMLSRYEKMEKIKQOHQMSDSISEGRLKKLQIPI 300
QY      334 EPKDDIHSLSQEEKELKRIQIDSSDPLSTEEKEFLKKLQIDIRDSLSEEEKELNRI 393
Db      301 EPKDDIHSLSQEEKELKRIQIDSSDPLSTEEKEFLKKLQIDIRDSLSEEEKELNRI 360
QY      394 QVDSNPULSEKEKEFLKKLQIDIPYDINQRLQDTGGLIDSPSINLDVRYKQYKRDIQND 453
Db      361 QVDSNPULSEKEKEFLKKLQIDIPYDINQRLQDTGGLIDSPSINLDVRYKQYKRDIQND 420
QY      454 ALLHOSISTLYNKIYLENNMINNL/TATLGADLVSDNTNFKINGINFEKKNFKYSIS 513
Db      421 ALLHOSISTLYNKIYLENNMINNL/TATLGADLVSDNTNFKINGINFEKKNFKYSIS 480
QY      514 SNYMTVDINERPALDNERLKWRIQSPTRAGYLENGKLILORNIGLEIKDVOIIOKSEK 573
Db      481 SNYMTVDINERPALDNERLKWRIQSPTRAGYLENGKLILORNIGLEIKDVOIIOKSEK 540
QY      574 EYIRIDAVVPKSKIDTQIOEAQLINOEWNKALGLPKYTKLITFNVHNRVNASNIVESAY 633
Db      541 EYIRIDAVVPKSKIDTQIOEAQLINOEWNKALGLPKYTKLITFNVHNRVNASNIVESAY 600
QY      634 LILNEMKNKIQSDLIKKTNTNYLDGNGRFVFTDITLPNIAEQYTHODEIYEVHSGKLYV 693

```


TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-8

Query Match 31.6%; Score 1309; DB 1; Length 472;
Best Local Similarity 98.8%; Pred. No. 9,4e-74;
Matches 255; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 34 AGGHDGVMHYKKEKKDKDEERNTQOEHLKIMKHIVKIEVKEEAVKKEAAE 93
|
DB 4 AGGHDGVMHYKKEKKDKDEERNTQOEHLKIMKHIVKIEVKEEAVKKEAAE 63
QY 94 KLEKVPDVLVEMKKAIGKIYVDGDTKHSLEALSEDKKIKIDYGDALLHEHYV 153
|
DB 64 KLEKVPDVLVEMKKAIGKIYVDGDTKHSLEALSEDKKIKIDYGDALLHEHYV 123
QY 154 AKESYEPVLVIOSEEDYVENTEKALNYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 213
|
DB 124 AKESYEPVLVIOSEEDYVENTEKALNYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 183
QY 214 SDSGDDLLFTNOLKEHPTDFSVFELPONSNEVOEFAKAFAYITEPOHNDVLOLYAPEA 273
|
DB 184 SDSGDDLLFTNOLKEHPTDFSVFELPONSNEVOEFAKAFAYITEPOHNDVLOLYAPEA 243
QY 274 FNYMDKFNQOEINLSLEE 291
|
DB 244 FNYMDKFNQOEINLTRAEE 261

RESULT 5
US-08-082-849B-8
Sequence 8, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Kimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-8

Query Match 31.6%; Score 1309; DB 1; Length 472;
Best Local Similarity 98.8%; Pred. No. 9,4e-74;
Matches 255; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 34 AGGHDGVMHYKKEKKDKDEERNTQOEHLKIMKHIVKIEVKEEAVKKEAAE 93
|
DB 4 AGGHDGVMHYKKEKKDKDEERNTQOEHLKIMKHIVKIEVKEEAVKKEAAE 63
QY 94 KLEKVPDVLVEMKKAIGKIYVDGDTKHSLEALSEDKKIKIDYGDALLHEHYV 153
|
DB 64 KLEKVPDVLVEMKKAIGKIYVDGDTKHSLEALSEDKKIKIDYGDALLHEHYV 123
QY 154 AKESYEPVLVIOSEEDYVENTEKALNYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 213
|
DB 124 AKESYEPVLVIOSEEDYVENTEKALNYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 183
QY 214 SDSGDDLLFTNOLKEHPTDFSVFELPONSNEVOEFAKAFAYITEPOHNDVLOLYAPEA 273
|
DB 184 SDSGDDLLFTNOLKEHPTDFSVFELPONSNEVOEFAKAFAYITEPOHNDVLOLYAPEA 243
QY 274 FNYMDKFNQOEINLSLEE 291
|
DB 244 FNYMDKFNQOEINLTRAEE 261

RESULT 6
PCT-US94-01624-8
Sequence 8, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Kimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Steuart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043

; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-01624-8

Query Match 31.6%; Score 1309; DB 5; Length 472;
Best Local Similarity 98.8%; Pred. No. 9,4e-74;
Matches 255; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 34 AGHGDSVGHVHVEKKNKDEERKKTQEHKKEIMKHIVKTEVGEAVKKEAAE 93
DB 4 AGHGDSVGHVHVEKKNKDEERKKTQEHKKEIMKHIVKTEVGEAVKKEAAE 63
OY 94 KLEKVPDYLEMYKAIGGIYIVDGDITKHSLSLSESDKKIKINDYKGDALLHHEHYV 153
DB 64 KLEKVPDYLEMYKAIGGIYIVDGDITKHSLSLSESDKKIKINDYKGDALLHHEHYV 123
OY 154 AKEGYEPVLIQSSSEYVENTEKALNYYEIGKILSRDLSKINOPYOKFLDVLNTIKNA 213
DB 124 AKEGYEPVLIQSSSEYVENTEKALNYYEIGKILSRDLSKINOPYOKFLDVLNTIKNA 183
OY 214 SPSDGDLLFTNQLKEHPTDFSVFLEQNSNEVOEVFAKAFAYIIEPQHRDVLQLYAPEA 273
DB 184 SPSDGDLLFTNQLKEHPTDFSVFLEQNSNEVOEVFAKAFAYIIEPQHRDVLQLYAPEA 243
OY 274 FNYMDKFNEQEIINLSLE 291
DB 244 FNYMDKFNEQEIINLTRA 261

RESULT 7
US-08-021-601-10
; Sequence 10, Application US/08021601
; Patent No. 5591631

; GENERAL INFORMATION:

; APPLICANT: Leppla, Stephen H.
; APPLICANT: Kilmpel, Kurt R.
; APPLICANT: Nichols, Peter J.
; APPLICANT: Atora, Naveen
; APPLICANT: Singh, Yogendra
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021.601
; FILING DATE: 19930212
; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880

; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 amino acids

; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-021-601-10

Query Match 31.5%; Score 1307; DB 1; Length 508;
Best Local Similarity 99.6%; Pred. No. 1,4e-73;
Matches 254; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 34 AGHGDSVGHVHVEKKNKDEERKKTQEHKKEIMKHIVKTEVGEAVKKEAAE 93
DB 1 AGHGDSVGHVHVEKKNKDEERKKTQEHKKEIMKHIVKTEVGEAVKKEAAE 60
OY 94 KLEKVPDYLEMYKAIGGIYIVDGDITKHSLSLSESDKKIKINDYKGDALLHHEHYV 153
DB 61 KLEKVPDYLEMYKAIGGIYIVDGDITKHSLSLSESDKKIKINDYKGDALLHHEHYV 120
OY 154 AKEGYEPVLIQSSSEYVENTEKALNYYEIGKILSRDLSKINOPYOKFLDVLNTIKNA 213
DB 121 AKEGYEPVLIQSSSEYVENTEKALNYYEIGKILSRDLSKINOPYOKFLDVLNTIKNA 180
OY 214 SPSDGDLLFTNQLKEHPTDFSVFLEQNSNEVOEVFAKAFAYIIEPQHRDVLQLYAPEA 273
DB 181 SPSDGDLLFTNQLKEHPTDFSVFLEQNSNEVOEVFAKAFAYIIEPQHRDVLQLYAPEA 240
OY 274 FNYMDKFNEQEIINLS 288
DB 241 FNYMDKFNEQEIINLT 255

RESULT 8
US-08-082-849B-10
; Sequence 10, Application US/08082849B
; Patent No. 5677274

; GENERAL INFORMATION:

; APPLICANT: Leppla, Stephen H.
; APPLICANT: Kilmpel, Kurt R.
; APPLICANT: Atora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/082,849B
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/021,601
; FILING DATE: 12-FEB-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-161-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-10

Query Match 31.5%; Score 1307; DB 1; Length 508;
Best Local Similarity 99.6%; Pred. No. 1.4e-73;
Matches 254; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGGHDVGMHVKKEKKNDEKRNKRDEERNKTOEHLKELIMKHYIEVKGEEAVKKEAAE 93
|||||
DB 1 AGGHODVGMHVKKEKKNDEKRNKRDEERNKTOEHLKELIMKHYIEVKGEEAVKKEAAE 60
QY 94 KLEKVPSPDLVEMKKAIGKIIYVDGDTTKHISLEALSEDKKKIKIDYKDALHHEHYV 153
|||||
DB 61 KLEKVPSPDLVEMKKAIGKIIYVDGDTTKHISLEALSEDKKKIKIDYKDALHHEHYV 120
QY 154 AKEGEPLVLIQSSSEDIYENTKALNYYEIGKILSRILSKINOPYOKFLDVLNTTKNA 213
|||||
DB 121 AKEGEPLVLIQSSSEDIYENTKALNYYEIGKILSRILSKINOPYOKFLDVLNTTKNA 180
QY 214 SDSGODLLFTNOLKEHPTDSVEFLQNSNEVOEVFAKAFAYIIEPQHRDVLQYAPEA 273
|||||
DB 181 SDSGODLLFTNOLKEHPTDSVEFLQNSNEVOEVFAKAFAYIIEPQHRDVLQYAPEA 240
QY 274 FNYMDKFNEOEINLS 288
|||||
DB 241 FNYMDKFNEOEINLT 255

RESULT 9
PCT-US94-01624-10
Sequence 10, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND KHOURI and CREW
STREET: Stewart Street Tower, 20th floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 508 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

PCT-US94-01624-10

Query Match 31.5%; Score 1307; DB 5; Length 508;
Best Local Similarity 99.6%; Pred. No. 1.4e-73;
Matches 254; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGGHDVGMHVKKEKKNDEKRNKRDEERNKTOEHLKELIMKHYIEVKGEEAVKKEAAE 93
|||||
DB 1 AGGHODVGMHVKKEKKNDEKRNKRDEERNKTOEHLKELIMKHYIEVKGEEAVKKEAAE 60
QY 94 KLEKVPSPDLVEMKKAIGKIIYVDGDTTKHISLEALSEDKKKIKIDYKDALHHEHYV 153
|||||
DB 61 KLEKVPSPDLVEMKKAIGKIIYVDGDTTKHISLEALSEDKKKIKIDYKDALHHEHYV 120
QY 154 AKEGEPLVLIQSSSEDIYENTKALNYYEIGKILSRILSKINOPYOKFLDVLNTTKNA 213
|||||
DB 121 AKEGEPLVLIQSSSEDIYENTKALNYYEIGKILSRILSKINOPYOKFLDVLNTTKNA 180
QY 214 SDSGODLLFTNOLKEHPTDSVEFLQNSNEVOEVFAKAFAYIIEPQHRDVLQYAPEA 273
|||||
DB 181 SDSGODLLFTNOLKEHPTDSVEFLQNSNEVOEVFAKAFAYIIEPQHRDVLQYAPEA 240
QY 274 FNYMDKFNEOEINLS 288
|||||
DB 241 FNYMDKFNEOEINLT 255

RESULT 10
US-08-021-601-6
US-08-021-601-6
Sequence 6, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P. C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021.601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-9880
TELEFAX: 404/688-0770
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-6

Query Match 31.5%; Score 1306; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 1,4e-73;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGGHDVGAMHVKKEKKNKDENRKDEERKKTQOEHLKEIMKHIVKEVGEAVKKEAAE 93
DB 1 AGGHDVGAMHVKKEKKNKDENRKDEERKKTQOEHLKEIMKHIVKEVGEAVKKEAAE 60
QY 94 KLEKVPSPVLEMYKAIGKITYVGDITKHTLSLALSDDKKIKIDYKGDALLHEHHY 153
DB 61 KLEKVPSPVLEMYKAIGKITYVGDITKHTLSLALSDDKKIKIDYKGDALLHEHHY 120
QY 154 AKEGEYPLVIOSSSEDYVENTEKALNVYIEIGKILSRDLSKINOPYKFLDVLNTIKNA 213
DB 121 AKEGEYPLVIOSSSEDYVENTEKALNVYIEIGKILSRDLSKINOPYKFLDVLNTIKNA 180
QY 214 SDSGODLLFTNQLKEHPTDFSVFLEQNSNEVOEFAKAFAYYIEPQHRDVLQIYAPEA 273
DB 181 SDSGODLLFTNQLKEHPTDFSVFLEQNSNEVOEFAKAFAYYIEPQHRDVLQIYAPEA 240
QY 274 FNYMDKFNEOEINL 287
DB 241 FNYMDKFNEOEINL 254

RESULT 11

US-08-082-849B-6
Sequence 6, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arota, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-6

Query Match 31.5%; Score 1306; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 1,4e-73;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGGHDVGAMHVKKEKKNKDENRKDEERKKTQOEHLKEIMKHIVKEVGEAVKKEAAE 93
DB 1 AGGHDVGAMHVKKEKKNKDENRKDEERKKTQOEHLKEIMKHIVKEVGEAVKKEAAE 60
QY 94 KLEKVPSPVLEMYKAIGKITYVGDITKHTLSLALSDDKKIKIDYKGDALLHEHHY 153
DB 61 KLEKVPSPVLEMYKAIGKITYVGDITKHTLSLALSDDKKIKIDYKGDALLHEHHY 120
QY 154 AKEGEYPLVIOSSSEDYVENTEKALNVYIEIGKILSRDLSKINOPYKFLDVLNTIKNA 213
DB 121 AKEGEYPLVIOSSSEDYVENTEKALNVYIEIGKILSRDLSKINOPYKFLDVLNTIKNA 180
QY 214 SDSGODLLFTNQLKEHPTDFSVFLEQNSNEVOEFAKAFAYYIEPQHRDVLQIYAPEA 273
DB 181 SDSGODLLFTNQLKEHPTDFSVFLEQNSNEVOEFAKAFAYYIEPQHRDVLQIYAPEA 240
QY 274 FNYMDKFNEOEINL 287
DB 241 FNYMDKFNEOEINL 254

RESULT 12

PCT-US94-01624-6
Sequence 6, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arota, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Steuart Street Tower, 20th Floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-6

Query Match 31.5%; Score 1306; DB 5; Length 456;
Best Local Similarity 100.0%; Pred. No. 1,4e-73;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

Db 1063 LMKLENISSTEGVETVEHEVQNVYVDVDPAMKQDFGLINEAGLKMFPNLEDPVK 1122
Qy 212 NMSDSGODLLTNNQKEPTDFSVFLEQNSNEVOEFAKAFAYIEPQHRDVLQVAP 271
Db 1123 SES-----DVITVEEIKDEP-----VOKEVEKETVSIIEEMEEIVDVLBE 1163
Qy 272 EAFNMYDKF-----NEDEINLSLEELKQORLSREKEKEKIK 308
Db 1164 EREDLTKMIDAVESEIETSSDSKETESIKQEKQVSLVEEVDONDESEVLEK 1223
Qy 309 OHYOHV-----SDLSSEGRGL-----LKKQIPRE-----PKKD 338
Db 1224 NMEELMKDAVEINDITSLIEETQELNEVADLIKDMEKLEKALBEDSKELIADAD 1283
Qy 339 DIHLSIOBEKEL-----LKRQIDS-----SDFLSTEE-----KEF----- 370
Db 1284 DYLEKVEIEEHDTITLDEVELEKQVEEDKIEKVSPLKDEEDILKEVEIKELESEILE 1343
Qy 371 -----LKKLOIDIRDSL-----EKEKELNRIQVDSNPSEKKEFLKRLKLDI 416
Db 1344 DYKELKTITDLEEKKELEKQHEKFESEAEIKDLEADILKEVSSLEVEEKKLE--- 1400
Qy 417 QPYDINQRLQD-----TGGLIDSPSINDVRKQYKRDQIONIDA 454
Db 1401 EYHELKEEVEHIIISGDANIKGLEEDLEEVDDLKSGIILMLKGMDELMGMDKSELDVTT 1460
Qy 455 LHOISGSLYKTYILEMMNINLTATGADLVSTDMTKINRG-----IFNEKKNF 508
Db 1461 KGEREESLAK-----DVLSSALGMDDEQMKTRKKAQRPKLEVLLEKEVEKEP 1508
Qy 509 KXSISNNYIVDINERPALD-----NEBLKMRIO--LSPPTRGYLENGCLLIQORNI 558
Db 1509 KKKTKKKKRFPIKDEPEDEVEEMKDEDEIEDEVEDIEEDKEDKVEDIDEDIDEDI 1568
Qy 559 GLEIKD--VOIKOSEKEXIRIDAKVVPKSKIDTKIOEQLNINQEMNALGPKYTKLI 616
Db 1569 G-EDDEVIDLIVQKEKRIEKYKAK---KKLEKKEVEEGVSGLKKHIVDEVM---KVVQKI 1621
Qy 617 TPNVHNRYASNIVESAYLILEMKNKNIQSDLIKKTNYL 655
Db 1622 DKEV-DKEVSKALESKNDVTNVLKON--ODPFSKYKNEFV 1657

RESULT 15
; Sequence 6, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-604A-6

```

```

Query Match 5.0%; Score 208; DB 1; Length 2710;
Best Local Similarity 20.4%; Pred. No. 0.00011;
Matches 196; Conservative 139; Mismatches 277; Indels 348; Gaps 51;

Qy 48 EKKKDKNRKDEERKNTQOEHLKET-MKHIVKTEVKG--EEVYKKEAAEKLLEK---V 99
Db 212 EYNRDETVLESYRTNS-----LKKINSNHGIDRANSLFTEQDLNIYSELELRGNLAA 266
Qy 100 PSDVLEM--YKATGKIYI-----VDDGITKHS-----LEASBKKK 136
Db 267 ASDIVRLALAKNFG--VYLDVDMLPQIHSDLEFTISRPSISGLDRRDMKIKLEAIKRYKY 325
Qy 137 IKDIYKDALLHEHYVAYKEGEPV-----LVYQSEEDYVENTERALN----- 179
Db 326 INN-----YTSNFPKLDQOLKDNFKLIESEKSEKSEIFSKLENLNVSDLEI 372
Qy 180 -VYIEIGKILSRDILK-----INOPYQKFLDVLNTIKNASDSGODLLFTNOLK- 228
Db 373 KIAFALGSVINQALISKOSYLTNLVIEQYKNRYQFLNGLHNPALIESDNN--FTDTTGI 429
Qy 229 EHPDTSVERLEQNSNEVOEFAKAFAYIE-----POHRDVLQVAPFAV--NYMDKPNF 282
Db 430 FHDLSLFNSATAENS-----MFLTKTAPYLQVGFMEARSTJLSGGAASAYYDFIND 483
Qy 283 OEINL-----SLEELKQORLSRYE-----RWKIKQHYQMSDS 317
Db 484 QENTIKETLKASDLIEFKFPENNLSQLTGEINLSMSFQASAKYQFEKVRDYT--GGS 541
Qy 318 LSEEG-----RGLKLIKQIP-----IEPKKDI-----IHS 343
Db 542 LSEDNQVDNPKNTALDKNVLNNKIPSNVVEAGSKNYHYITQLOGDDISYEATCNLPS 601
Qy 344 LSOEKKELKRIQIDS--SDFLSTEEKFLKQIDIRDSLESEKELLN----- 391
Db 602 KNPKNSTIIOQRNNNESAKSFUSDGSESTLELKNYRIPKELNKKKRVKVTFTIGHGDEFN 661
Qy 392 -----RIQVDSNPSEKKEFLKRLKLDIOP-----YDINQRLQDYGGLI 432
Db 662 TSEFARLSYDS--LSNETSSFLDTIKLDISPRKNVENVNLGCMFSDYDNVEETVYCKLL 718

```

```

OY 433 DS-----PSINDV----RKQYKRDION----IDALLH-----OSIGSTLYNKI 468
Db 719 LSINDKITSTLPDVKNKSITIGANQYEVRIINSEGRKELLASHGKWINKEAIAIMSDLSKE 778
OY 469 YLYENMNINNLATLGLADLVSTDNFKINRGIFNEFKKNFKYSISSNYM--IYDINERPA 526
Db 779 YIF-----FDSIDN-----KLAKSKNIPGLASISE--- 804
OY 527 LDNERLKWRIQLSPDTRAGYLENGKLLIQRNIGLEIKDVQIIOSEKEYIRID---AKV 582
Db 805 -DIKTLLLDASVSPDKF-IINNLKLNIESSIG-----DIYIEKLEPYKN 848
OY 583 VPKSKIDTKIOEAOI--NINQEMNKALGLPKYTKLITFNVHNRYSNIVESAYLILNEMK 640
Db 849 IHNISIDLIDEFNILEWVSD-----LYELKRL-----NMLDEKYLISFE-- 889
OY 641 NNIOSDLIKKVTNYLV---DNGRFEVFTDITLPIIAEQYTHODEIYEQVHSGKGLYPES 696
Db 890 -----DISKNNSTYSVRPINKSNQESVIVE-TEKEIFSKYS-EHITKEIST----- 933
OY 697 RSLILHGSPSKGVELRNDSEGFIEFGHAVDDYAGYLLDKNOSDLVT-----NSKKFIDIF 751
Db 934 -----IKNS-----IITDVNGNLLDNQLDHTSQVNTLNAAPFIQSL 970
OY 752 KEEGSN--LTSYGRTEAEFFFAFRLMST--DHAERLKVQKNAPKTFQFINDQIKFI 806
Db 971 IDYSSNKDVNLNDLSTSVKQVLYAQLFTGTLNTIYDSIQLVNLISNA-----VNDTINVL 1024

```

Search completed: December 2, 2001, 13:49:10
 Job time: 190 sec

[illegible]

PA (MATECUN A J.
XX Galloway DR, Mateczun AJ;
XX WPI: 2001-408540/43.
DR N-PSDB: AAC86015.
XX
XX Protecting animal against lethal infection with Bacillus anthracis, by
PT administering wild type or mutated form of Bacillus anthracis lethal
PT factor protein or its fragment or a nucleic acid encoding the mutated
PT protein -
PS
PS Claim 3; Fig 1; 33pp: English.
XX
XX This sequence represents the B. anthracis lethal factor (LF). An
CC immunogenic fragment of LF, LF4, can be used to produce an immune
CC response which protects an animal against lethal injection with
CC Bacillus anthracis. DNA encoding the B. anthracis LF can be used
CC in conjunction with DNA encoding the protective antigen (PA) in a
CC DNA vaccine. Using a DNA vaccine which encodes the mutated LF protein
CC or fragment alone or in combination with a DNA encoding the PA protein
CC or its fragment, both components (humoral and cell-mediated) of the
CC immune system are stimulated, which results in longer term immune
CC memory response. The combined use of a mutated LF and PA gene or their
CC fragments results in a higher level of immune response, as judged by
CC overall serum antibody titers for LF and PA antigens, than the use of
CC either LF or PA genes in separate immunizations.
SQ
SQ Sequence 809 AA:

Query Match 100.0%; Score 4145; DB 22; Length 809;
Best Local Similarity 100.0%; Pred. No. 2.9e-245;
Matches 809; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNIKKEFIVISMSCLVTAITLSPVFIPLVVGAGHGVGMHVKEKKNKDKRDEE 60
DB 1 mnikkefivismsclvtaitlspvfiprlvgagghgvgmhvkeknkdkndekidee 60
QY 61 RNKTOEHLKETMKNIIVKIEVKGEEAVKKEAEKLEKVPDVLNMYKAIIGKRIYVGD 120
DB 61 rnktoehleketmkniiivkievkggeavkkeaekllekvpdvlennykaiigkriyvkd 120
QY 61 rnktoehleketmkniiivkievkggeavkkeaekllekvpdvlennykaiigkriyvkd 120
DB 61 rnktoehleketmkniiivkievkggeavkkeaekllekvpdvlennykaiigkriyvkd 120
QY 121 IRKHTSLESLSDKKIKRIYKGDALHHEHYVYAKGEPVLYTOSSEYVENTEKALV 180
DB 121 irkhtsleslsdkkikriykgdallhehyvayakgepvlytossedyventekalv 180
QY 121 irkhtsleslsdkkikriykgdallhehyvayakgepvlytossedyventekalv 180
DB 121 irkhtsleslsdkkikriykgdallhehyvayakgepvlytossedyventekalv 180
QY 181 YVEIGKILSRDLISKINQPYQKFLDVLNLIKNASDSGDLLFTNOLKHPHDFSEFLE 240
DB 181 yveigkilsrdliskinqpyqkflvlnliknasdsdgdllftnolkhphdfsefle 240
QY 181 yveigkilsrdliskinqpyqkflvlnliknasdsdgdllftnolkhphdfsefle 240
DB 181 yveigkilsrdliskinqpyqkflvlnliknasdsdgdllftnolkhphdfsefle 240
QY 241 QNSNEVQEVFAKAFAYIERPQHRDVLQLYAPEAFNTMDKFNEOEINLSLEELKQRMISR 300
DB 241 qnsnevqevfakafayierpqrhrdvlylapeafntmdkfneoeinlsleelkqrmisar 300
QY 241 qnsnevqevfakafayierpqrhrdvlylapeafntmdkfneoeinlsleelkqrmisar 300
DB 241 qnsnevqevfakafayierpqrhrdvlylapeafntmdkfneoeinlsleelkqrmisar 300
QY 301 YKWEKIKOHYHMSDSISEEGGILKTIQIPIPEPKKDDIHSISEEKLKRIQIDSS 360
DB 301 ykwekikohyhmssdsiseeggilktqipiepkddihsisgeekllkriqidss 360
QY 301 ykwekikohyhmssdsiseeggilktqipiepkddihsisgeekllkriqidss 360
DB 301 ykwekikohyhmssdsiseeggilktqipiepkddihsisgeekllkriqidss 360
QY 361 DELSTEENEKFLKQLQIDIRDSISEEKEELNRIQVDSNPISKEKEFEFKTKLIDQPD 420
DB 361 delsteenekflkqlqidirdsiseeekelnriqvdsnpiskekefektklidxpd 420
QY 361 delsteenekflkqlqidirdsiseeekelnriqvdsnpiskekefektklidxpd 420
DB 361 delsteenekflkqlqidirdsiseeekelnriqvdsnpiskekefektklidxpd 420
QY 421 INQRIQDTGGLIDSPSINIDVRKQYKRDQIQLDALLHQSISGLVINKIYLVENMINNTL 480
DB 421 inqriqdtgglidpsindvrkqykrdqilqldallhqsisgltvinkiylevenminntl 480
QY 421 inqriqdtgglidpsindvrkqykrdqilqldallhqsisgltvinkiylevenminntl 480
DB 421 inqriqdtgglidpsindvrkqykrdqilqldallhqsisgltvinkiylevenminntl 480
QY 481 ATLGADLVSTNTKINRGIFNEPKKFKKYSISSNMIYDINERPLDNERLQKRWOLSP 540
DB 481 atlgadlvstntkinrgifnepkkfkkysssnmiydinerpldnerlqkwrqlsp 540
QY 481 atlgadlvstntkinrgifnepkkfkkysssnmiydinerpldnerlqkwrqlsp 540
DB 481 atlgadlvstntkinrgifnepkkfkkysssnmiydinerpldnerlqkwrqlsp 540
QY 541 DTRAGTLENGKILQINILKTDVQIIKQSEKEYIRIDAKVVPKSKIDTKIQEADLNIN 600
DB 541 dtragtlenkglilqinilktvqiiqosekeyiridakvvpkskidtkiqeadi 600

DB 541 dtragtlenkglilqinilktvqiiqosekeyiridakvvpkskidtkiqeadi 600
QY 601 OEWNKALGLPKYTKLITTFVNHNRYSNIVESAYLILNEKKNINQSDILIKVTNYLVDGNG 660
DB 601 oewnkalgpkyltklittfvnhnraysnivesaylilnekkninqsdiliktvtnylvdgng 660
QY 661 RYVFTDITLPTNAEOYTHODEIYGVHSGGLVPESSRSLILHGSPKGVRLRNDSECFIHE 720
DB 661 ryvftditlptnaeoythodeiygvhsgglvpe SSRSLILHGSPKGVRLRNDSECFIHE 720
QY 721 FGHAVDDYAGYLLDKNQSDLVNNSKKFIDIPKEEGSNLTSYGRTEAEFFAEFLMHST 780
DB 721 fghavddyagylldknqsdlvnnskkfidipkeegsnltsygrteaeffae flmhst 780
QY 781 DHAERLKVONAKPTPOFINDQIKFIINS 809
DB 781 dhaerlkvonakptpofindqikfiins 809

RESULT 2
AAR60178
ID AAR60178 standard; Protein; 776 AA.
XX
XX AAR60178;
AC AAR60178;
XX 03-APR-1995 (first entry)
DT
XX
XX Lethal factor of Bacillus anthracis.
DE
XX
XX Anthrax: Bacillus anthracis; fusion protein; lethal factor;
KW protective antigen; cell killing; targeting; targeting; pathogen;
KW Intracellular; HIV; human immunodeficiency virus; toxin.
XX
XX Bacillus anthracis.
OS
XX
XX W09418332-A.
PN
XX
XX 18-AUG-1994.
PD
XX
XX 14-FEB-1994; 94WO-0501624.
PF
XX
XX 12-FEB-1993; 93US-0021601.
PR
XX 25-JUN-1993; 93US-0082849.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA
XX
XX Aroza N, Klimpel K, Leppia SH, Nichols PJ, Singh Y;
PI WPI: 1994-279753/34.
DR N-PSDB: AAQ70179.
XX
XX Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumour cells
PT or HIV-infected cells
PT
XX
XX
PS Disclosure; Page 75-77; 124pp: English.
XX
XX The sequence encoding the lethal factor of Bacillus anthracis may be
CC used in the construction of a nucleic acid which encodes a fusion
CC protein comprising the anthrax protective antigen binding domain of
CC the native anthrax lethal factor and a sequence encoding an activity
CC inducing domain of a second protein. The fusion proteins are useful
CC for the specific killing of tumour cells or the killing of cells
CC infected with intracellular pathogens, especially HIV.
XX
XX
SQ Sequence 776 AA:

Query Match 96.1%; Score 3983; DB 15; Length 776;
Best Local Similarity 100.0%; Pred. No. 2.2e-235;
Matches 776; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGGHGVGMHVKEKKNKDKRDEEHLKIMKIIIVIEVKGEEAVKKEAE 93

```

Db      1 agghgdvgmhvkexkdenkrkdeernktgeehlkemlkhvkvkgeavkkaeae 60
QY      94 KLEEKVPSDVLWEMKKAIGGKIYVDDGITHISLEALSEKSKKIKIDYGCDAALLHENYV 153
Db      61 kilekypsdvlemykaigkilyvdgdltkhislealsekdkkikldiygkdallhenyvy 120
QY      154 AKEGYEPVLVIOSESDYVENTEKALANYYEIGKILSRDILSKINOPORFLDVLNTTKNA 213
Db      121 akegyepvlyiqsedvyentekalnvyeygkilstdlisklnqpyqkfldvlnltkna 180
QY      214 SDSGDQLFTNLQKHPDTFSVEFLQNSNEQVEWFAKAFAYIIEPQHRDVLQLYAPEA 273
Db      181 sdsdgdqlftnlqkhpdtfsveflqnsneqvewfakafayyiepqhrdvqllyapea 240
QY      274 FNYMDFNQEDELSTLEELKQDMLSRYEKWEKIKOHYQHSWSLSSEEGGLKLDQIPT 333
Db      241 fnymdkfneqelnlsleelkqdmisryekwekikqhyqhsdsisseegrgllkklqipl 300
QY      334 EPKKDITISLSQEEKELKRIQIDSSDFLSTBEKEFLKLDIDIRDSLSEEEKELNRI 393
Db      301 epkddilnlsqeekeflkrlqidsdflstbekeflkrlqidsisseekellnrl 360
QY      394 QVDSNPLSEKEKEFLKRLKLDIQPYDINQRLQDTGGLDPSINLDVRKQYKRDIONID 453
Db      361 qvdsnplsekekeflkrlkldiqpydinqrlqdtgglidspsindvrkqykrldiqnid 420
QY      454 ALLHOSIGSTLYNKIYLYEMKMNINNLATLIGALVDSDTNKRIGFNFKNFKYSIS 513
Db      421 allhgisgstlynykilyemmnlnlatalga dlvdstdtclnrglfnefkknfkysis 480
QY      514 SNMYAIDINERPALDNERLKWRIQDSPTRAGYLENGKLLQNLNIGELIRDOVQIKQSEK 573
Db      481 snymyidinerpaldnlerlkwriqispdtragylengklllqnlnglrelkdvqklyqsek 540
QY      574 EYRIDAKVVPKSKIDTKIQEAOLNINQEWNKALGLPKYTKLITFVNHVNRVASYVESAY 633
Db      541 eyridakvvpkskikdtkqeqalninqewnkalglpkytkliltfvnhvnrvasyvesay 600
QY      634 LILNEKKNINOSDLIKKYNVYLVDSNGRPFYDTITPLNIEQYTHODEIYQVHSKLYV 693
Db      601 lllneknlnqsdllkklvnylvdsngrfvfdltlplnaeqylhdeqlyeqvhskslyv 660
QY      694 PESRSILHGPSKGVELRNDSEGFIEHFGAVDVGYLDDXNOSDLYVTSKKPFIDIFKE 753
Db      661 peersllhgpskgvelrndsegfiehfgnavodagylldkxnsdlvtskpfidilfke 720
QY      754 EGSNLTSGRTNDEEFAEAFRLMHSTDAERLKVONAPKTFQFINDQIKFIINS 809
Db      721 egsltsygrtneaeffaefrlmhstdaerlkvgnapktftqfinqdilkfiins 776

RESULT      3
AAU00222
ID      AAU00222 standard; Protein: 485 AA.
AC      AAU00222;
XX
DT      31-MAY-2001 (first entry)
XX
DE      Lfn-Bcl-XL apoptosis-modifying fusion protein.
XX
KW      Human: Lfn-Bcl-XL; apoptosis: cancer; spinal muscular atrophy;
KW      anthrax lethal factor; neoplasm; tumor; hyper-proliferation;
KW      Alzheimer's disease; neurodegenerative disorder; stroke;
KW      transient ischaemic neuronal injury; spinal cord injury;
KW      Huntington's disease.
XX
XX      Chimeric - Homo sapiens.
OS      Chimeric - Corynebacterium diptheriae.
OS      Chimeric - Synthetic.
XX
FH      key      Location/Qualifiers

```

```

FT      Region      5...10
FT      /note="6x histidine tag"
FT      Region      21..276
FT      /note="Anthrax lethal factor amino acids 1 to 255"
FT      Region      277..485
FT      /note="Bcl-XL amino acids 1 to 209"
XX
PN      W0200112661-A2.
XX
PD      22-FEB-2001.
XX
PF      15-AUG-2000; 2000MO-US22293.
XX
PR      16-AUG-1999; 99US-0149220.
XX
PA      (HARD ) HARVARD COLLEGE.
PA      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI      Youle RJ, Liu X, Collier RJ;
XX
DR      WPI; 2001-218343/22.
DR      N-PSDB; AAS00250.
XX
PT      Novel fusion protein for modifying apoptosis in target cell and
PT      reducing apoptosis after transient ischaemic neuronal injury, has two
PT      domains which targets protein to a cell and modifies apoptotic response
PT      of cell
XX
PS      Claim 4; Page 64-65; 65pp: English.
XX
CC      The sequence represents the amino acid sequence of Lfn-Bcl-XL apoptosis-
CC      modifying fusion protein comprising anthrax lethal factor (LF) sequence
CC      fused to Bcl-XL. The functional apoptosis-modifying fusion protein is
CC      capable of binding a target cell and integrating into or crossing a
CC      cellular membrane of the target cell. The apoptosis-modifying fusion
CC      protein comprises at least two domains: the DFR domain, which targets
CC      the fusion protein to the target cell and the Bcl-XL domain, which
CC      modifies an apoptotic response of the target cell. The fusion protein is
CC      useful for modifying (inhibiting or enhancing) apoptosis in a target
CC      cell, such as neuron, lymphocyte, cancer, neoplasm, macrophage.
CC      epithelial, stem, tumor or hyper-proliferative cell or an adipocyte. It
CC      is also useful for reducing apoptosis in a subject after transient
CC      ischaemic neuronal injury, especially spinal cord injury. The fusion
CC      protein may be used to treat various diseases and injury conditions
CC      through inhibition or enhancement of apoptotic cellular response,
CC      including neurodegenerative disorders such as Alzheimer's disease,
CC      Huntington's disease, spinal muscular atrophy, stroke episodes and
CC      unregulated cell growth as in tumors and various cancers. The apoptosis-
CC      modifying fusion protein can be delivered effectively throughout the body
CC      and targeted to selective tissue and cells.
XX
SQ      Sequence      485 AA:

Query Match      31.9%; Score 1322; DB 22; Length 485;
Best Local Similarity 58.8%; Pred. No. 4,7e-73;
Matches 293; Conservative 30; Mismatches 49; Indels 126; Gaps 11;

```


CC	components.
XX	
SQ	Sequence 508 AA;

Query Match	31.5%	Score 1307;	DB 15;	Length 508;
Best Local Similarity	99.6%	Pred. No. 4.1e-72;		
Matches 254; Conservative	1;	Mismatches 0;	Indels 0;	Gaps 0

Qy	34	AGGAGDGMNHYKKEKKNDEKRRDEERNKTOEHLKEIMKHIVKIEVKEEVAKKEAAE	93
Db	1	aggnvdrvgmhyvkekehndekrrkdeernktgeehlkelnkhvklevgyeeavkkeaee	60
Qy	94	KLEKRVSDVLEWYKAIGKRIYVDGITKHSLEALSDEKKIKIDYIGDALLHENYU	153
Db	61	kllkxvpsdvlewykalgkriyivdgdltkhslealsedkkikidylgxdallhenyuy	120
Qy	154	AKGEYEPVLYGSESDVENMERKALNYYEELGKILISDITKINOPVKRDLVNLTKNA	213
Db	121	akegyepvlyvgssedvyenkealnuyyegkllisdlstklnpqykrdlvnltkna	180
Qy	214	SDSGODLLEFTNOLKEHPDTSVEFLLEBONSNEVOEYAKAFAYITEFQHRDVLQILAPEA	273
Db	181	sdsdgqdlflfnqlkehpdtsweflleqnsnewevfakafayitepqhrdvlqilyapea	240
Qy	274	FNYMDKREDEEINLS	288
Db	241	fnymdkrfnedeinlt	255

RESULT 6

AAR60180 standard; Protein; 456 AA.

AC AAR60180;

DT 04-APR-1995 (first entry)

DE LF(1-254) --TR--PE(401-602) toxin fusion protein.

KM Anthrax *Bacillus anthracis*; fusion protein; lethal factor;
 KM Protective antigen; cell killing; targeting; pathogen
 KM Intracellular; HIV; human immunodeficiency virus; toxin;
 KM *Pseudomonas*; exotoxin.
 XX
 OS *Bacillus anthracis*.
 OS *Pseudomonas* sp.

PN W09418332-A.

PD 18-AUG-1994.

PF 14-FEB-1994; 94WO-US01624.

PR 12-FEB-1993; 93US-0021601.

PR 25-JUN-1993; 93US-0082849.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Arora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;

DR WPI; 1994-279753/34.

DR N-PSDB; AAR6

PT Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumour cells
PT or HIV-infected cells

PS Claim 7; Page 86-87; 124pp; English

This sequence is a fusion protein comprising amino acid residues 1-254 of the anthrax protective antigen binding domain of the native anthrax lethal factor, a two residue linker and residues

CC 401-602 of a Pseudomonas exotoxin A activity inducing domain of a
CC second protein. Such toxin fusion proteins may be useful for the
CC specific killing of tumour cells or the killing of cells infected
CC with intracellular pathogens, especially HIV, depending on their
CC components.

Sequence 456 AA;

Query Match	31.5%	Score 1306;	DB 15;	Length 456;
Best Local	100.0%	Pred. No.	4.1e-72;	
Matches 254;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	34	AGGHGDVGMHNYKKEKEXKKNDEKRRKDRBENRKTQGEHLKEIMKNHYLIEVKEEBAVKKAAE	93
Db	1	agghgdvgmhnykkekexkknndekrrkdrbenrktqgeehlkeimkhyli evkveeaakkkaae	60
QY	94	KLLEKVPSPDVLMEWKATIGKTYIYDGDITKHSLEAKSEDKKKTKTDYGRDALLHENVY	153
Db	61	kllkvpstvdlmwykatsgkytyivdgditkhtslsalseddkkkktdygrdaallhenyy	120
QY	154	AKEGEPEPLVYQSSDDYVENETKALNYEYELGKLLSRDILSKINOPYQKGLDVLTNTKNA	213
Db	121	akegepeplvlyqssddyvenetekalnuyeyelgkllsrdlisrklndpyqkftldvntlnkna	180
QY	214	SDSDODDLFFNOLKEHPHTDSVFLEQNSNEQVFEAKAFAYITGEQNHBDVLOLYAPEA	273
Db	181	sdsdoddlffnqlkehphtdsvfleqnsnevgevfakafayitgeqnhbdvlyllyapea	240
QY	274	FNYYDKFNEQELNL 287	
Db	241	fnymdkfnegeinl 254	

RESULT 7

ID AAR04236 standard; protein; 800 AA.

AC AAR04236;

DT 12-SEP-1989 .(first entry)

DE Adenyl cyclase gene of *Bacillus anthracis*.

OS	<i>Bacillus anthracis</i> .
XX	
KM	adenyl cyclase; pertussis; protective vaccines; signal sequence
...	

PN EP366550-A.

PD 02-MAY-1990.

PF 25-OCT-1989; 89EP-0402949.

PR 25-OCT-1988; 88FR-0013952.

PA (INSP) INSTITUT PASTEUR.

PI Escuyer V, Duflot E, Mock M, Danchin A;

DR WPI; 1990-133988/18.

DR N-NSDB; Q04123

PT Nucleotide sequence encoding adenyl cyclase of *Bacillus anthracis*
PT and derived proteins, useful in protective vaccines, also effective
PT against pertussis

PS Claim 8; 13; 23pp; French.

In vivo the adenylyl cyclase protein is synthesised as a precursor with a signal sequence. The mature protein is secreted into the periplasmic space, the signal peptide having been cleaved off at the moment of secretion.


```
Db 377 -----lkvremdlekrehnflh-medqklklnsfvkn--nqlkvykce--- 418
QY 160 PVLIOSSDYVENTEKALNVYVEIGKILSRDLSKINPYOKFLDYLVNTIKNASDSQG 219
Db 419 ----lknltkekekelk---dienvskeeklnlqnekeqjlafnkhke--- 468
QY 220 DLLFTNQLKEHPTDFSVFELEONSNEVOEYFAKAFAYTIEPQHRDYLOLYAPAFYVMDK 279
Db 469 ----lhgleelke-svktltietgelgem-----vdikkelqdl-----gek 507
QY 280 FNEQELNLSLEBKDOQMLSRV-----EKWEKIKOHYOHWSLSLEE----- 321
Db 508 ynaqlesielskkekkeyngkntyjeelnlnkeleentkynlqnnynneimln 567
QY 322 ----GSGLKIKQIPIEPKKDOI-----IHSOSEKEL---LKKIQIDSSPLSTEEK 368
Db 568 dlhmlngalktmntqstlknvdhllneqdlknnkgtlnsktlnselvngjmdl--keek 625
QY 369 EFLKRIQIDIRDSLESEKELNRIQVDSSNPLSEKEKEFLKRLK----- 414
Db 626 dflnqjvqlnql-----dltrkmeekenkmllegenkykgemellrgnksesnlm 680
QY 415 DIQPIDINORL-----QDTGGLIDSPSINLDVRKQYKRDIONIDALLHQS 459
Db 681 deevcdlkxklksesemkmmkeehkklaelkdcdvr--tremeknedklnmlkee- 737
QY 460 IGSITLXNKIYLVENNMNINLTATLGADIVDSTDNTR-IRGRINERKKNKYSISS--- 514
Db 738 -----yed-kin-----tlkeqnedklnltlkeqnedklnltkeyekhlnlmkee 781
QY 515 -NYMIVDINERPALD-----NERLKNRIQISPTIRAGYL-----ENGR 551
Db 782 yehklnltlnegnekhlnltlnegnekhlnlmkeeyedkmmtnlneqnedkmmksleeyenk 841
QY 552 LILORNGILEIDY--QIIKOSKEKIRIDAKVVPKSKIDTKIOEQLNINO----- 601
Db 842 nqlnsnneklkldvnyneyleevdklvtldek---kkqfdkelnyahlkakeheqjllte 898
QY 602 -EMNKALGPKYTKLITFNVHNRYASNIYESAVLIIN-EMKNNIOS-DLIKKTNTLVDS 658
Db 899 meelkqrdnkysdl-----yekyl-kllkslcmllnleecddlenedlrrleeyinn 952
QY 659 NGRFVETDITLPIAIOYTHODEIYEQVHSGLYVPSRSILHGPSKGEVLNDESGFI 718
Db 953 kgl-----kveeekenhk-----hsfnllkskxf- 979
QY 719 HEPGAHVADYAGYLDKNOSDLVTNSKR-----IDIFKEGGSNL 758
Db 980 --fksniedkshelkkkkekdlkskdeeknkklkelndklklqdellykq-sna 1036
QY 759 TSGVRNEAEFFAEAFRLMHSSTDAERLKYOKNAPRTFOFINOIKFI 806
Db 1037 qqvdhkkkswlllkdkskelkxkenglnvexneekdlkkkddelfil 1084

RESULT 9
AAM39097
ID AAM39097 standard; Protein; 2663 AA.
XX
AC AAM39097;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2242.
XX
KW Human; nootropic; immunosuppressant; cytosstatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
```

```
XX
PN MO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000MO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RJ;
XX
DR WPI: 2001-442253/47.
DR N-PSDB; AA158253.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 4; SEQ ID NO 2242; 10078pp; English.
XX
SS The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AA42213) with nootropic,
CC immunosuppressant and cyostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokine activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 2663 AA:
XX
Query Match 5.7%; Score 235.5; DB 22; Length 2663;
Best Local Similarity 20.5%; Pred. No. 1e-05;
Matches 178; Conservative 155; Mismatches 305; Indels 231; Gaps 41;
XX
QY 42 MHVKEKKNK-----DENKRRKDEERNKQF-EHLKEIKKH-----YKREV 81
Db 524 melkkekndldefealerltkkgemqilheislknlykhhevyngdlenelsskvel 563
QY 82 KGEBAVKKAEAEKULEKVPDLEMYKAIGKITIYVDGITKHSLEALSDEKRRKIK-- 138
Db 584 lre---kedqklkqeyidsqklienlkm-----dis--yslesi-edbpqmkqtl 627
QY 139 -----DIYGDALLHEHYVAKGEYEVLYIQSSEDEVVENTEKALNYYEFGKILSR 190
Db 628 fdaetvaldakresafitsenlelkekme--latykgemndiqlygsqleakkmqy 684
QY 191 DISIKIQOPQKFLDYLVNTIKNASDSG---QDLFTNQLKEHPTDFSVF---LEONSN 244
Db 685 dlekeqsaenleltsll-----dqkypkdlllnlelegklttdqkqlnveeneea 738
QY 245 EVOGEVF---AKAFAYTIEPQHRDV-----LOLYAPEAFYVMDK---NEQELNLS 288
Db 739 lreevlllseklpseverlrlkelqdkseelhltse---xdklifsevnhesrvqgl 794
```

Oy	289	LEEL--KDOMLSRYEKMEKRIQOHQWMSDLSSEBGRLLKLQIPRPPKDDIHS--	343
Db	795	leejgkttdd-lattgsnykstqdeiqnf-----kllhmdfeqykmyleene	841
Oy	344	-LSQEEKELKRRIQ-IDSS-DFLSTE-----EKEFEKLQIDIRD	380
Db	842	rmngeaylnlskaakffssjgalktelsygtelgqektrevgerlnemeglike-qlemrđ	900
Oy	381	S---LSEEEKELLRIOVDSSNPLSEKEKEFEKLKLIDQPYDINORLODT-----	428
Db	901	spjqaverektl-----ltekjqtleevktlcqekddklqjesiqierdqk	949
Oy	429	GGIDSPSINIDVAKQYKRRIQNIIDALHOSISTLTKNYILEXENNINNLATIGADLV	488
Db	950	sdlnltvmmldldeqqltnaleslqg-hqetlnltkskiseevsnln-----me	998
Oy	489	DSTONTKINRGIFNEFFKNKPKYSISSNYMIVDINERPALDNERLKWRIQSPDRAGYLE	548
Db	999	entgetk-----defgk-----mvgldkkqgdleakn-----	10253
Oy	549	NGKLILORNICLEIKDVOIIRKQSEKEYIRIDAKVVPKSIIDFKIOE-AQLNINDEMNKL	607
Db	1026	-----tqltadvkđneelieqgkifsljqlqeknelqumlesviaeqkltcdlken---	1076
Oy	608	GLPKYTKLITFNVNRRASNVESAVYLINEMKNINQSOILKVTNYLYVDNGRVPFDI	667
Db	1077	-----lemtlenq-----eellrljlgelak-k-qgelvqeknhalkkegelstřcd	1120
Oy	668	TLPNIAEQTYHQDDIVQVHSKGLYVPESRSLILHGPRSGVELRNDSEGFIEFHAAYDV	727
Db	1121	rlaeeveeklleksgqqlqekgqqllnvqeemseqkklnelnlknelkketllehmete	1180
Oy	728	--YAGYLLDKKQSDL--VTNSKFKI---DIFEESGSNLTSTGRTNEAEFFAEAFRL---	776
Db	1181	rlqlaqklnenyeevksalkterkrylkelqsksfeterdhrlgyrlrelaatqqltkeelkia	1240
Oy	777	-MHS7DHAERLKVOKN--APKPTGFINDQ	802
Db	1241	hnhlkehqetldelrřvsektqalntq	1269
RESULT 10			
AAM40883			
ID	AAM40883 standard; Protein; 2688 AA.		
AC	AAM40883;		
XX			
DT	22-OCT-2001 (first entry)		
XX			
DE	Human polypeptide SEQ ID NO 5814.		
XX			
KW	Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	autochthonic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokineitic; thrombolytic; drug screening; arthritits; inflammation;		
KW	leukaemia.		
XX			
OS	Homo sapiens.		
XX			
PN	MO20015312-A1.		
XX			
PD	26-JUL-2001.		
XX			
PF	26-DEC-2000; 2000MO-U0534263.		
XX			
PR	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		

PR		19-OCT-2000; 2000US-0693036.
PR		29-NOV-2000; 2000US-0727344.
xx		
PA	(HYSE-) HYSEQ INC.	
Pt	Tang YF, Liu C, Asundi T, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
Pt	Wang Z, Wang Z, Wehrman V, Xu C, Xue AJ, Yang Y, Zhang J;	
Pt	Zhao QJ, Zhou P, Goodrich R, Dimanac RT;	
xx		
DR	WPI: 2001-442253/47.	
xx	N-PsDB: AA160039.	
Pt	Novel nucleic acids and polypeptides, useful for treating disorders	
Pt	such as central nervous system injuries -	
PS	Example 2: SEQ ID NO 5814; 10078pp; English.	
xx		
CC	The invention relates to human nucleic acids (AA157798-AA161369) and	
CC	the encoded polypeptides (AAM36642-AAM42213) with nootropic,	
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful	
CC	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	of the invention may be used to treat diseases of the peripheral nervous	
CC	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	localised neuropathies and central nervous system diseases, such as	
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	utilisation of the activities such as: Immune system suppression,	
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic	
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening	
CC	assays for receptor activity, arthritis and inflammation, leukaemia and	
CC	C.N.S disorders.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification.	
xx		
SQ	Sequence 2688 AA:	
	Query Match 5.7%; Score 235.5; DB 22; Length 2688;	
	Best Local Similarity 20.5%; Pred. No. 1e-05; Mismatches 305; Indels 231; Caps 41	
	Matches 178; Conservative 155;	
OY	42 MHVEKEKKK-----DENKRKDEERNRTOE-EHLKEIKKH-----IVKIEV 81	
	: : : : : :	
Db	548 melklkxkndldfealelexttkkkdgmqljheislknlkylkhreyngqdenelsakvel 607	
OY	82 KEEBAVKKAEAEXLEKVSDVLDMTKAIGKTIYVDGTTIKITISLASEDKKKR-- 138	
	: : : : : : :	
Db	608 lxe-----kedqikllgelyidsqklenkm-----dis--yslsli-edbpkmgktl 651	
OY	139 -----DIYGDALHHEHYVAKGGEYEVLIOSSEDVENKTEKALNANYELGKILSR 190	
	: : : : : : : : : : : : :	
Db	652 fdaetvaldakresalisenlelkeknke---latcykqmendidqlygsqleaakkmyv 708	
OY	191 DILSKINQYOKFLDVLNTIKNASDSG---ODILETNOLEKPTDFSVF---LEQNSN 244	
	: : : : : : : : : : : : :	
Db	709 dlekelqsfneltktstl -----dgvpkallcnlenlegitdlqkelnveeneaa 762	
OY	245 EVQEYF---AKAFAYIEPQRHV-----IQLYAPEAFNYDKF-----NEQETINS 288	
	: : : : : : : : : : : :	
Db	763 lreevlllselslpseverltkeiqdksesehliltse----dklifsevvhkstrvgjl 818	
OY	289 LPEEL---KPQRMLSRKEKMEKITQHVOHMSDSLSEGGRGLKKTLOIPIEPKODIHNS-- 343	
	: : : : : : : : : : : :	
Db	819 ljeigtktkd-latqsnkykstcdqeqntf-----kcllmfdqekyxmwleene 865	
OY	344 -LSOEKELLKRIQ-IDSS-DPLSTE-----EKFEKLKLTIIDIRD 380	
	: : : : : : : : : : : : : :	
Db	866 rmngelvnlskaeqktdsslgalktelisyktglqgektrevgerlnemeqllke-qletrnd 924	
OY	381 S---LSEEKELLNRIOVSSNPLSKKEKEFFLKLLKDLOIPYIDINORLODT----- 428	
	: : : : : : : : : : : : :	
Db	925 splqtverekt ------itekltqgtleevenfltekdskdlqyslterdqkl 973	

Matches 162; Conservative 136; Mismatches 271; Indels 248; Gaps 37;

```

QY 45 KEKEKN-----KDENRKDEERKKTQEHLEIM-----KHIVKIEVGEAVKKE 90
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 677 kdlskndnlnkdynlvqkkskkkkkflndlnltnfteskygdlyvkggees--ke 734
QY 91 AAEKLEKV-----PSDYLEMYKAIGKITYIVDPDITKIHSLSELSDEK 134
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 735 dknqldfvqecyrnndlrctnksdflknlkidnkkyei---ynleqegeinekk 791
QY 135 K-----KIKDIYGDALLHEHYVYAKEGEPVLYQSSSEDEVVENTEKALNV 180
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 792 nynknndsnkftfflkleneffkdkllldsgqifgdslladikenyrcadnldnnnksl 851
QY 181 YVEIGKLSLDLSKINQPYQKFLDVNTIKNASDS---DGQDLLEFTNOLKEHP--TDFS 235
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 852 yedgenfltr-----nepltneyeeknnliysdeqkyneediifkdkikekknnds 905
QY 236 VEFLEQNSNEVOEFAKAFAYVTEPQHRDVLQLYAPEAFNVMDKFNQEIINLS---LEE 291
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 906 sddfences--vge-----klyvnekieynknndkssssssilllee 945
QY 292 LKDQR-----MLSRYEKMEKIKQHYQHWS-----DSLSEEGGGLLK 328
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 946 lkykkekdelvsnlcvlldefehndlenyisvsddmktvsknntlgyvenkvdk 1005
QY 329 LQPIEPKKDDIHSLSQEKELKRIQIDSSDFLSTEKEFLKLOIDIRDSISEEKE 388
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1006 tnveydkkgddgyielsfeshkleeekfdanniynddeleknlk---kdyisdvdkn 1062
QY 389 LLNRIOVDSSNPSEKEFEFLKKLKDIOPIYDINQRLQDPTGGLDPSI---NLDVRKOY 445
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1063 hvnnl-ynlergedereenef-----venkigsceshksnefictenkslltkgy 1109
QY 446 --KRDIQNIIDLHQSIGSTLYNKIYLYENMNINMLTATLGAD-LVDSTDTNFKINRGIFN 502
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1110 mskedlsenvrll-----kddlnlnlskqnyfeallldk-----k 1142
QY 503 EFKKNFRYSISSNMVYDINERFALDNERLKWRIQLSPDRAGYLENGKLILORNIGLEI 562
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1143 qvndntfgmnieqn-----ndkIk-edkIdegayfeylednkll----- 1179
QY 563 KDVQIIRKQSEKEYIRIDAKVVPKSKIDTKIQEAOQUNINOEWNKAL-----GLPKYTKL 615
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1180 --dsylketnke-----neelikeyk---klkknleindemddiklllnfgyipyiqsp 1230
QY 616 ITFNVNHRVYASN-----IVESAYLLINEMKNKNIQSOLIKKYVNTYLVDSNGRFFFTDIT 668
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1231 cesaegcsylnknkycdaflsdsdvlvfsgkcvlknfnkkkt----- 1274
QY 669 LPNIAEQTQODELYEQ--VHSK-GLYVPESRSI--LLHGPS-----KGVELRNDSEGFTH 719
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1275 -----veveyekkaieekiglyqeelnlsllcgcdylgvnhgigivnale-llk 1322
QY 720 EFGHAVDDYAGYLLDKNQSDLVNTSKKEFID--IFKEE 754
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1323 afnufed-----lklkldivsnpftkldkmynee 1352

```

Search completed: December 2, 2001, 13:48:11
Job time: 131 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:49:10 ; Search time 55.64 Seconds
(without alignments)
314.658 Million cell updates/sec

Title: US-09-747-521-2_COPY_1_778
Perfect score: 3987
Sequence: 1 MNKKKFIKTVISMSCLVTAL.....TSYGRTEAEFFAEAFRLMH 778

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3825	95.9	776	1 US-08-021-601-2	Sequence 2, Appl
2	3825	95.9	776	1 US-08-082-849B-2	Sequence 2, Appl
3	3825	95.9	776	5 PCT-US94-01624-2	Sequence 2, Appl
4	1309	33.8	472	1 US-08-021-601-8	Sequence 8, Appl
5	1309	33.8	472	1 US-08-082-849B-8	Sequence 8, Appl
6	1309	33.8	472	5 PCT-US94-01624-8	Sequence 8, Appl
7	1307	32.8	508	1 US-08-021-601-10	Sequence 10, Appl
8	1307	32.8	508	1 US-08-082-849B-10	Sequence 10, Appl
9	1307	32.8	508	5 PCT-US94-01624-10	Sequence 10, Appl
10	1306	32.8	456	1 US-08-021-601-6	Sequence 6, Appl
11	1306	32.8	456	1 US-08-082-849B-6	Sequence 6, Appl
12	1306	32.8	456	5 PCT-US94-01624-6	Sequence 6, Appl
13	501	12.6	800	6 5183745-3	Patent No. 5183745
14	228.5	5.7	1786	4 US-08-973-462-8	Sequence 8, Appl
15	202	5.1	2710	1 US-08-480-604A-6	Sequence 6, Appl
16	202	5.1	2710	4 US-08-405-496A-6	Sequence 6, Appl
17	202	5.1	2710	2 US-08-915-136-6	Sequence 6, Appl
18	194.5	4.9	3248	1 US-08-353-700-1	Sequence 1, Appl
19	194.5	4.9	3248	5 PCT-US95-16216-1	Sequence 1, Appl
20	191.5	4.8	1164	4 US-08-923-992A-2	Sequence 2, Appl
21	188	4.7	976	4 US-09-104-324B-4	Sequence 4, Appl
22	186.5	4.7	1164	4 US-08-923-992A-10	Sequence 10, Appl
23	186	4.7	1588	5 PCT-US93-07261-11	Sequence 11, Appl
24	186	4.7	1663	5 PCT-US93-07261-16	Sequence 16, Appl
25	186	4.7	2285	4 US-09-308-375-2	Sequence 2, Appl
26	184.5	4.6	1388	2 US-08-685-576-4	Sequence 4, Appl
27	184	4.6	990	2 US-08-645-193B-15	Sequence 15, Appl

28	182	4.6	2482	1 US-08-328-254-6	Sequence 6, Appl
29	181	4.5	1312	2 US-08-592-126-148	Sequence 148, App
30	181	4.5	1312	2 US-08-687-080-51	Sequence 51, Appl
31	180.5	4.5	1128	4 US-08-923-992A-6	Sequence 6, Appl
32	180	4.5	990	2 US-08-392-625-20	Sequence 20, Appl
33	180	4.5	990	2 US-08-466-961A-20	Sequence 8, Appl
34	178.5	4.5	1098	4 US-08-923-992A-8	Sequence 1, Appl
35	176.5	4.4	1388	2 US-08-685-576-1	Sequence 1, Appl
36	172.5	4.3	1073	4 US-09-541-782-6	Sequence 6, Appl
37	171	4.3	1104	4 US-08-923-992A-4	Sequence 4, Appl
38	169	4.2	872	1 US-08-766-014-2	Sequence 4, Appl
39	168.5	4.2	1038	4 US-09-541-782-4	Sequence 4, Appl
40	165	4.1	663	4 US-09-196-293-5	Sequence 5, Appl
41	165	4.1	663	4 US-08-209-603B-5	Sequence 5, Appl
42	164.5	4.1	1057	4 US-09-541-782-10	Sequence 10, Appl
43	163.5	4.1	1311	2 US-08-460-309-4	Sequence 4, Appl
44	163.5	4.1	3111	2 US-08-125-077-4	Sequence 4, Appl
45	163	4.1	663	4 US-08-235-836C-78	Sequence 78, Appl

ALIGNMENTS

RESULT 1
US-08-021-601-2
; Sequence 2, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Nichols, Peter J.
; APPLICANT: Atoria, Naveen
; APPLICANT: Singh, Yogendra
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021, 601
; FILING DATE: 19930212
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36, 016
; REFERENCE/DOCKET NUMBER: 1414, 057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-9880
; TELEFAX: 404/688-0770
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-021-601-2

Query Match 95.9%; Score 3825; DB 1; Length 776;
Best Local Similarity 100.0%; Pred. No. 3.5e-239;
Matches 745; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
34 AGGAGVGVMHVKKKKKNDKRRKNTQEHKLKIKHIVKTEVGEAEVKKRAAE 93

```
Db 1 AGGHGADVAMHAKKKEKKNDEKRNKQOEHLKTIKMHYKIEVKGGEAVKKEAAE 60
QY 94 KLEKVPDVLDMYKAIGKIYIYDGDITTKHISLEALSEDKRKIKDIYKGDALLHEHYV 153
Db 61 KLEKVPDVLDMYKAIGKIYIYDGDITTKHISLEALSEDKRKIKDIYKGDALLHEHYV 120
QY 154 AKEGEPELVYIOSSSDYVENTEKALNYYEIGKILSRDLSKINOPYOKFLDVLNTIKNA 213
Db 121 AKEGEPELVYIOSSSDYVENTEKALNYYEIGKILSRDLSKINOPYOKFLDVLNTIKNA 180
QY 214 SDSGQODLLFTNOLKEHPTDVSVEFLKQNSNVEOVFAKAFAYIIEPOHRYVLOLYAPEA 273
Db 181 SDSGQODLLFTNOLKEHPTDVSVEFLKQNSNVEOVFAKAFAYIIEPOHRYVLOLYAPEA 240
QY 274 FNYMDKFNQEIINLSLEELKQDRLMSRYEKWEKIKQHYQHSNDSISEGRGLKKLOIPI 333
Db 241 FNYMDKFNQEIINLSLEELKQDRLMSRYEKWEKIKQHYQHSNDSISEGRGLKKLOIPI 300
QY 334 EPKDDIITHLSQOEKELKRIQIDSSDFLSTEEKEFLKKIQIDIRDSLSEEEKELNRI 393
Db 301 EPKDDIITHLSQOEKELKRIQIDSSDFLSTEEKEFLKKIQIDIRDSLSEEEKELNRI 360
QY 394 QVDSNPNLSEKEKEFLKKLKDIOPIYDINORLODTGGLIDSPSINLDVRYKQYKRDIONID 453
Db 361 QVDSNPNLSEKEKEFLKKLKDIOPIYDINORLODTGGLIDSPSINLDVRYKQYKRDIONID 420
QY 454 ALLHOSIGSTLYNKIYLYENNINNL7ATIGADVSTDNTRKINRGIFNEKKNEKYSIS 513
Db 421 ALLHOSIGSTLYNKIYLYENNINNL7ATIGADVSTDNTRKINRGIFNEKKNEKYSIS 480
QY 514 SNMYIVDINERPALDNEELKRIQISPTPRAGYLENGKLLIQRNIGLEIKVOIITKQSEK 573
Db 481 SNMYIVDINERPALDNEELKRIQISPTPRAGYLENGKLLIQRNIGLEIKVOIITKQSEK 540
QY 574 EYIRIDAKVVPKSKIDTQIOEAQNLINQEWKALGLPYTKLITFNVRASNIIVEASAY 633
Db 541 EYIRIDAKVVPKSKIDTQIOEAQNLINQEWKALGLPYTKLITFNVRASNIIVEASAY 600
QY 634 LILNEMKNNIOSDLIKKTYNTLVDSNGRFVFTDITLPIAEOYTHODEIYEQVHSGKLYV 693
Db 601 LILNEMKNNIOSDLIKKTYNTLVDSNGRFVFTDITLPIAEOYTHODEIYEQVHSGKLYV 660
QY 694 PESNISILHGSQVGLNDSSEGFIERGHAVDPAVGYLLDKNOSDLVATNSKKFLDIFKE 753
Db 661 PESNISILHGSQVGLNDSSEGFIERGHAVDPAVGYLLDKNOSDLVATNSKKFLDIFKE 720
QY 754 EGSNLTSGRTNEAEFFAEFRLMH 778
Db 721 EGSNLTSGRTNEAEFFAEFRLMH 745

RESULT 2
US-08-082-849B-2
; Sequence 2, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Aroca, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
; TITLE OF INVENTION: Related Methods
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESS: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/082,849B
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/021,601
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-161-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-082-849B-2

Query Match 95.9%; Score 3825; DB 1; Length 776;
Best Local Similarity 100.0%; Pred. No. 3.5e-239;
Matches 745; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGGHGADVAMHAKKKEKKNDEKRNKQOEHLKTIKMHYKIEVKGGEAVKKEAAE 93
Db 1 AGGHGADVAMHAKKKEKKNDEKRNKQOEHLKTIKMHYKIEVKGGEAVKKEAAE 60
QY 94 KLEKVPDVLDMYKAIGKIYIYDGDITTKHISLEALSEDKRKIKDIYKGDALLHEHYV 153
Db 61 KLEKVPDVLDMYKAIGKIYIYDGDITTKHISLEALSEDKRKIKDIYKGDALLHEHYV 120
QY 154 AKEGEPELVYIOSSSDYVENTEKALNYYEIGKILSRDLSKINOPYOKFLDVLNTIKNA 213
Db 121 AKEGEPELVYIOSSSDYVENTEKALNYYEIGKILSRDLSKINOPYOKFLDVLNTIKNA 180
QY 214 SDSGQODLLFTNOLKEHPTDVSVEFLKQNSNVEOVFAKAFAYIIEPOHRYVLOLYAPEA 273
Db 181 SDSGQODLLFTNOLKEHPTDVSVEFLKQNSNVEOVFAKAFAYIIEPOHRYVLOLYAPEA 240
QY 274 FNYMDKFNQEIINLSLEELKQDRLMSRYEKWEKIKQHYQHSNDSISEGRGLKKLOIPI 333
Db 241 FNYMDKFNQEIINLSLEELKQDRLMSRYEKWEKIKQHYQHSNDSISEGRGLKKLOIPI 300
QY 334 EPKDDIITHLSQOEKELKRIQIDSSDFLSTEEKEFLKKIQIDIRDSLSEEEKELNRI 393
Db 301 EPKDDIITHLSQOEKELKRIQIDSSDFLSTEEKEFLKKIQIDIRDSLSEEEKELNRI 360
QY 394 QVDSNPNLSEKEKEFLKKLKDIOPIYDINORLODTGGLIDSPSINLDVRYKQYKRDIONID 453
Db 361 QVDSNPNLSEKEKEFLKKLKDIOPIYDINORLODTGGLIDSPSINLDVRYKQYKRDIONID 420
QY 454 ALLHOSIGSTLYNKIYLYENNINNL7ATIGADVSTDNTRKINRGIFNEKKNEKYSIS 513
Db 421 ALLHOSIGSTLYNKIYLYENNINNL7ATIGADVSTDNTRKINRGIFNEKKNEKYSIS 480
QY 514 SNMYIVDINERPALDNEELKRIQISPTPRAGYLENGKLLIQRNIGLEIKVOIITKQSEK 573
Db 481 SNMYIVDINERPALDNEELKRIQISPTPRAGYLENGKLLIQRNIGLEIKVOIITKQSEK 540
QY 574 EYIRIDAKVVPKSKIDTQIOEAQNLINQEWKALGLPYTKLITFNVRASNIIVEASAY 633
Db 541 EYIRIDAKVVPKSKIDTQIOEAQNLINQEWKALGLPYTKLITFNVRASNIIVEASAY 600
QY 634 LILNEMKNNIOSDLIKKTYNTLVDSNGRFVFTDITLPIAEOYTHODEIYEQVHSGKLYV 693
Db 601 LILNEMKNNIOSDLIKKTYNTLVDSNGRFVFTDITLPIAEOYTHODEIYEQVHSGKLYV 660
```

Db 601 L1LWKKNNI0SDLLIKKVTNYLVGNGRFVFTDITLPLNIAEQYTHODEIYEQVHSGKLVY 660
QY 694 PERSRSLILHGPKSGVELRNDSEGFIEFGHAYVDVAGYLLDKN0SLVYNSKKFIDIFKE 753
Db 661 PERSRSLILHGPKSGVELRNDSEGFIEFGHAYVDVAGYLLDKN0SLVYNSKKFIDIFKE 720
QY 754 EGSNLTSGRTNEAEFFAFARLHM 778
Db 721 EGSNLTSGRTNEAEFFAFARLHM 745

RESULT 3
PCT-US94-01624-2
; Sequence 2, Application PC/TUS9401624
; GENERAL INFORMATION:
; APPLICANT: Leppia, Stephen H.
; APPLICANT: Kilmpel, Kurt R.
; APPLICANT: Aroza, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREM
; STREET: Steuart Street Tower, 20th Floor, One Market
; STREET: Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01624
; FILING DATE: June 25, 1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31, 677
; REFERENCE/DOCKET NUMBER: 15280-115
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-01624-2

Query Match 95.9%; Score 3825; DB 5; Length 776;
Best Local Similarity 100.0%; Pred No. 3.5e-239;
Matches 745; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGGHGDMVGHVKEKKNKDENRKRKDENRKTOEHLKEIMKHIVKLEVGEEAVKKEAAE 93
Db 1 AGGHGDMVGHVKEKKNKDENRKRKDENRKTOEHLKEIMKHIVKLEVGEEAVKKEAAE 60
QY 94 KLEKVPSPVLEMYKAIIGKTIYVGDITKHISLEALSSEDKKIKIYIGKDALIHEHYV 153
Db 61 KLEKVPSPVLEMYKAIIGKTIYVGDITKHISLEALSSEDKKIKIYIGKDALIHEHYV 120
QY 154 AKEGEPVAVIOSSSEYVNTKALNVVYIEIGTILSRILSROLINQYOKFLDVLTNTIKNA 213
Db 121 AKEGEPVAVIOSSSEYVNTKALNVVYIEIGTILSRILSROLINQYOKFLDVLTNTIKNA 180
QY 214 SDSGODLLFTNOLKEHPTDFSVFLEQNSNEQVEVFAFAFAYIIPQRHVDYQLVAPRA 273

Db 181 SDSGODLLFTNOLKEHPTDFSVFLEQNSNEQVEVFAFAFAYIIPQRHVDYQLVAPRA 240
QY 274 ENYMDKFNEQEIINLSLEELKDQRMLSRYEKWEKIKOHYQMSDSLSEEGRLKIKLOIPI 333
Db 241 ENYMDKFNEQEIINLSLEELKDQRMLSRYEKWEKIKOHYQMSDSLSEEGRLKIKLOIPI 300
QY 334 EPRKDDIINLSQEEKELKRIODSSDFLSTEKEFKLKLQIDIDISLSEEEKELNRI 393
Db 301 EPRKDDIINLSQEEKELKRIODSSDFLSTEKEFKLKLQIDIDISLSEEEKELNRI 360
QY 394 QVDSNPLSEKEKEFLKLLKLDIQPYDINORLODTGGLDPSINLDVROKQYRDIONID 453
Db 361 QVDSNPLSEKEKEFLKLLKLDIQPYDINORLODTGGLDPSINLDVROKQYRDIONID 420
QY 454 ALHQSIGSTLYNKKIYLYENMNINNLATLGLADLVSTDTKINRGIFFNEFKKRFYSIS 513
Db 421 ALHQSIGSTLYNKKIYLYENMNINNLATLGLADLVSTDTKINRGIFFNEFKKRFYSIS 480
QY 514 SNMYIVDINERPALDNBRKWRIOISPDPBAGYLGKGLILORINIGLEIKVOIIOSEK 573
Db 481 SNMYIVDINERPALDNBRKWRIOISPDPBAGYLGKGLILORINIGLEIKVOIIOSEK 540
QY 574 EYIRIDAKVYPKSKIDTKIOEAOLINQEWKKAIGLPKYTKLITFNVHNRYASNIVESAY 633
Db 541 EYIRIDAKVYPKSKIDTKIOEAOLINQEWKKAIGLPKYTKLITFNVHNRYASNIVESAY 600
QY 634 L1LWKKNNI0SDLLIKKVTNYLVGNGRFVFTDITLPLNIAEQYTHODEIYEQVHSGKLVY 693
Db 601 L1LWKKNNI0SDLLIKKVTNYLVGNGRFVFTDITLPLNIAEQYTHODEIYEQVHSGKLVY 660
QY 694 PERSRSLILHGPKSGVELRNDSEGFIEFGHAYVDVAGYLLDKN0SLVYNSKKFIDIFKE 753
Db 661 PERSRSLILHGPKSGVELRNDSEGFIEFGHAYVDVAGYLLDKN0SLVYNSKKFIDIFKE 720
QY 754 EGSNLTSGRTNEAEFFAFARLHM 778
Db 721 EGSNLTSGRTNEAEFFAFARLHM 745

RESULT 4
US-08-021-601-8
; Sequence 8, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppia, Stephen H.
; APPLICANT: Kilmpel, Kurt R.
; APPLICANT: Aroza, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,601
; FILING DATE: 19930212
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36, 016
; REFERENCE/DOCKET NUMBER: 1414, 057

TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-8

Query Match 32.8%; Score 1309; DB 1; Length 472;
Best Local Similarity 98.8%; Pred. No. 4.1e-77;
Matches 255; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 34 AGGAGDGMHVKKEKKDKENKRDENRNTQOEHLKEIMKHIVKIVKGEAVKKEAAE 93
D 4 AGGAGDGMHVKKEKKDKENKRDENRNTQOEHLKEIMKHIVKIVKGEAVKKEAAE 63
QY 94 KLEKVPDVLVEMKKAIGKIYVDGDTKHSLEALSEDKKKIKIDYGDALLHEHYV 153
D 64 KLEKVPDVLVEMKKAIGKIYVDGDTKHSLEALSEDKKKIKIDYGDALLHEHYV 123
QY 154 AKEGEYVLVQSSDEVVENTERKALNYYELGKILSDILSKINOPYOKFLDVNTIKNA 213
D 124 AKEGEYVLVQSSDEVVENTERKALNYYELGKILSDILSKINOPYOKFLDVNTIKNA 183
QY 214 SDSGDDLFTNQKEHPDPSVFELEQNSNEVOEFAKAFAYIEQHNDVLDLYAPEA 273
D 184 SDSGDDLFTNQKEHPDPSVFELEQNSNEVOEFAKAFAYIEQHNDVLDLYAPEA 243
QY 274 FNYMDFNEOEINLSLE 291
D 244 FNYMDFNEOEINLTRA 261

RESULT 5

US-08-082-849B-8
Sequence 8, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-8

Query Match 32.8%; Score 1309; DB 1; Length 472;
Best Local Similarity 98.8%; Pred. No. 4.1e-77;
Matches 255; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 34 AGGAGDGMHVKKEKKDKENKRDENRNTQOEHLKEIMKHIVKIVKGEAVKKEAAE 93
D 4 AGGAGDGMHVKKEKKDKENKRDENRNTQOEHLKEIMKHIVKIVKGEAVKKEAAE 63
QY 94 KLEKVPDVLVEMKKAIGKIYVDGDTKHSLEALSEDKKKIKIDYGDALLHEHYV 153
D 64 KLEKVPDVLVEMKKAIGKIYVDGDTKHSLEALSEDKKKIKIDYGDALLHEHYV 123
QY 154 AKEGEYVLVQSSDEVVENTERKALNYYELGKILSDILSKINOPYOKFLDVNTIKNA 213
D 124 AKEGEYVLVQSSDEVVENTERKALNYYELGKILSDILSKINOPYOKFLDVNTIKNA 183
QY 214 SDSGDDLFTNQKEHPDPSVFELEQNSNEVOEFAKAFAYIEQHNDVLDLYAPEA 273
D 184 SDSGDDLFTNQKEHPDPSVFELEQNSNEVOEFAKAFAYIEQHNDVLDLYAPEA 243
QY 274 FNYMDFNEOEINLSLE 291
D 244 FNYMDFNEOEINLTRA 261

RESULT 6

REG-US94-01624-8
Sequence 8, Application PC/US9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Steuart Street Tower, 20th Floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PC/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-5040
TELEFAX: (415) 543-5043

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-8

Query Match 32.8%; Score 1309; DB 5; Length 472;
Best Local Similarity 98.8%; Pred. No. 4, 1e-77;
Matches 255; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 34 AGCGDVGAMHVEKEKKNKDEERNKTOEHLKEIMKHIVKIEVGGEAVKKEAAE 93
DB 4 AGCGDVGAMHVEKEKKNKDEERNKTOEHLKEIMKHIVKIEVGGEAVKKEAAE 63
QY 94 KLEKVPDYLEMYKAIGGIYIVDGIKHSLEALSDEKKIKINDYKDALHHEHYV 153
DB 64 KLEKVPDYLEMYKAIGGIYIVDGIKHSLEALSDEKKIKINDYKDALHHEHYV 123
QY 154 AREGYEPVLIQSSSEYVENTEKALNVYIEIGILSRDILSKINOPYOKFLDVLNTIKNA 213
DB 124 AREGYEPVLIQSSSEYVENTEKALNVYIEIGILSRDILSKINOPYOKFLDVLNTIKNA 183
QY 214 SPSDGDLLFTNQLKHPDVSVEFLQNSNEVQVFAFAFAYIIEPQHRDVLQLYAPEA 273
DB 184 SPSDGDLLFTNQLKHPDVSVEFLQNSNEVQVFAFAFAYIIEPQHRDVLQLYAPEA 243
QY 274 FNYMDKFNEQEIINLSLE 291
DB 244 FNYMDKFNEQEIINLSLE 261

RESULT 7

US-08-021-601-10
Sequence 10, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Ieppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414,057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-9880
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 508 amino acids

TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-10

Query Match 32.8%; Score 1307; DB 1; Length 508;
Best Local Similarity 99.6%; Pred. No. 6, 1e-77;
Matches 254; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGCGDVGAMHVEKEKKNKDEERNKTOEHLKEIMKHIVKIEVGGEAVKKEAAE 93
DB 1 AGCGDVGAMHVEKEKKNKDEERNKTOEHLKEIMKHIVKIEVGGEAVKKEAAE 60
QY 94 KLEKVPDYLEMYKAIGGIYIVDGIKHSLEALSDEKKIKINDYKDALHHEHYV 153
DB 61 KLEKVPDYLEMYKAIGGIYIVDGIKHSLEALSDEKKIKINDYKDALHHEHYV 120
QY 154 AREGYEPVLIQSSSEYVENTEKALNVYIEIGILSRDILSKINOPYOKFLDVLNTIKNA 213
DB 121 AREGYEPVLIQSSSEYVENTEKALNVYIEIGILSRDILSKINOPYOKFLDVLNTIKNA 180
QY 214 SPSDGDLLFTNQLKHPDVSVEFLQNSNEVQVFAFAFAYIIEPQHRDVLQLYAPEA 273
DB 181 SPSDGDLLFTNQLKHPDVSVEFLQNSNEVQVFAFAFAYIIEPQHRDVLQLYAPEA 240
QY 274 FNYMDKFNEQEIINLS 288
DB 241 FNYMDKFNEQEIINLS 255

RESULT 8

US-08-082-849B-10
Sequence 10, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Ieppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 508 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-10

Query Match 32.8%; Score 1307; DB 1; Length 508;
Best Local Similarity 99.6%; Pred. No. 6,1e-77;

Matches 254; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGHGDVGMHVKKEKKNDEKRNKDEERNKTOEHLKEIMKHIVIEVKGEEAVKKEAAE 93
DB 1 AGHGDVGMHVKKEKKNDEKRNKDEERNKTOEHLKEIMKHIVIEVKGEEAVKKEAAE 60
QY 94 KLEKVPDVLMEYKAIGKIYVDGDTKHISLEALSDESKKIKIDYKDALLHEHYV 153
DB 61 KLEKVPDVLMEYKAIGKIYVDGDTKHISLEALSDESKKIKIDYKDALLHEHYV 120
QY 154 AKEGEPLVLIQSSSEDDYVENTEKALNVYEEIGKILSRDILSKINOPYKFLDVLNTIKNA 213
DB 121 AKEGEPLVLIQSSSEDDYVENTEKALNVYEEIGKILSRDILSKINOPYKFLDVLNTIKNA 180
QY 214 SDSGQDLLFTNQLKEHPTDVSVEFLQNSNEVOEVFAKAFAYIIEPQHRDVLQYAPEA 273
DB 181 SDSGQDLLFTNQLKEHPTDVSVEFLQNSNEVOEVFAKAFAYIIEPQHRDVLQYAPEA 240
QY 274 FNYMDKFNQOEINLS 288
DB 241 FNYMDKFNQOEINLT 255

RESULT 9

PCT-US94-01624-10
Sequence 10, Application PC/TUS9401624

GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
RELATED METHODS
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE and CREM
STREET: Stewart Street Tower, 20th Floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 508 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

PCT-US94-01624-10

Query Match 32.8%; Score 1307; DB 5; Length 508;
Best Local Similarity 99.6%; Pred. No. 6,1e-77;

Matches 254; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGHGDVGMHVKKEKKNDEKRNKDEERNKTOEHLKEIMKHIVIEVKGEEAVKKEAAE 93
DB 1 AGHGDVGMHVKKEKKNDEKRNKDEERNKTOEHLKEIMKHIVIEVKGEEAVKKEAAE 60
QY 94 KLEKVPDVLMEYKAIGKIYVDGDTKHISLEALSDESKKIKIDYKDALLHEHYV 153
DB 61 KLEKVPDVLMEYKAIGKIYVDGDTKHISLEALSDESKKIKIDYKDALLHEHYV 120
QY 154 AKEGEPLVLIQSSSEDDYVENTEKALNVYEEIGKILSRDILSKINOPYKFLDVLNTIKNA 213
DB 121 AKEGEPLVLIQSSSEDDYVENTEKALNVYEEIGKILSRDILSKINOPYKFLDVLNTIKNA 180
QY 214 SDSGQDLLFTNQLKEHPTDVSVEFLQNSNEVOEVFAKAFAYIIEPQHRDVLQYAPEA 273
DB 181 SDSGQDLLFTNQLKEHPTDVSVEFLQNSNEVOEVFAKAFAYIIEPQHRDVLQYAPEA 240
QY 274 FNYMDKFNQOEINLS 288
DB 241 FNYMDKFNQOEINLT 255

RESULT 10

US-08-021-601-6
Sequence 6, Application US/08021601

PATENT No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 12
RELATED METHODS
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spiralt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414,057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-6

Query Match 32.8%; Score 1306; DB 1; Length 456;
 Best Local Similarity 100.0%; Pred. No. 6.2e-77;
 Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGCHGVGMHVKERKKNKDEBNKTOEHLKEIKKHIVKIEVGEEAVKKEAAE 93
 DB 1 AGCHGVGMHVKERKKNKDEBNKTOEHLKEIKKHIVKIEVGEEAVKKEAAE 60

QY 94 KLEKPSVLEMYKAIGKITYVDGDIKHSLEALSEDKKIKIDYIGKDALHHEHYV 153
 DB 61 KLEKPSVLEMYKAIGKITYVDGDIKHSLEALSEDKKIKIDYIGKDALHHEHYV 120

QY 154 AKEGYEVLYIOSSSEYVENTEKALNVYIEIGKILSRDISKINOPYOKFLDVLNTIKNA 213
 DB 121 AKEGYEVLYIOSSSEYVENTEKALNVYIEIGKILSRDISKINOPYOKFLDVLNTIKNA 180

QY 214 SDSGDGLFTNOLKEHPDVSVEFLQNSNEVOEYFAKAFAYIIEPQHRDVLQLYAPEA 273
 DB 181 SDSGDGLFTNOLKEHPDVSVEFLQNSNEVOEYFAKAFAYIIEPQHRDVLQLYAPEA 240

QY 274 FNYMDKFNBOEINL 287
 DB 241 FNYMDKFNBOEINL 254

RESULT 11
 US-08-082-849B-6
 ; Sequence 6, Application US/08082849B
 ; Patent No. 5677274
 ; GENERAL INFORMATION:
 ; APPLICANT: Leppla, Stephen H.
 ; APPLICANT: Klimpel, Kurt R.
 ; APPLICANT: Atora, Naveen
 ; APPLICANT: Singh, Yogendra
 ; APPLICANT: Nichols, Peter J.
 ; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
 ; NUMBER OF SEQUENCES: 35
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, Eighth Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/082,849B
 ; FILING DATE: 25-JUN-1993
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/021,601
 ; FILING DATE: 12-FEB-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Weber, Kenneth A.
 ; REGISTRATION NUMBER: 31,677
 ; REFERENCE/DOCKET NUMBER: 15280-161-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 576-0200
 ; TELEFAX: (415) 576-0300
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 456 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-082-849B-6

Query Match 32.8%; Score 1306; DB 1; Length 456;
 Best Local Similarity 100.0%; Pred. No. 6.2e-77;
 Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGCHGVGMHVKERKKNKDEBNKTOEHLKEIKKHIVKIEVGEEAVKKEAAE 93
 DB 1 AGCHGVGMHVKERKKNKDEBNKTOEHLKEIKKHIVKIEVGEEAVKKEAAE 60

QY 94 KLEKPSVLEMYKAIGKITYVDGDIKHSLEALSEDKKIKIDYIGKDALHHEHYV 153
 DB 61 KLEKPSVLEMYKAIGKITYVDGDIKHSLEALSEDKKIKIDYIGKDALHHEHYV 120

QY 154 AKEGYEVLYIOSSSEYVENTEKALNVYIEIGKILSRDISKINOPYOKFLDVLNTIKNA 213
 DB 121 AKEGYEVLYIOSSSEYVENTEKALNVYIEIGKILSRDISKINOPYOKFLDVLNTIKNA 180

QY 214 SDSGDGLFTNOLKEHPDVSVEFLQNSNEVOEYFAKAFAYIIEPQHRDVLQLYAPEA 273
 DB 181 SDSGDGLFTNOLKEHPDVSVEFLQNSNEVOEYFAKAFAYIIEPQHRDVLQLYAPEA 240

QY 274 FNYMDKFNBOEINL 287
 DB 241 FNYMDKFNBOEINL 254

RESULT 12
 PCT-US94-01624-6
 ; Sequence 6, Application PC/TUS9401624
 ; GENERAL INFORMATION:
 ; APPLICANT: Leppla, Stephen H.
 ; APPLICANT: Klimpel, Kurt R.
 ; APPLICANT: Atora, Naveen
 ; APPLICANT: Singh, Yogendra
 ; APPLICANT: Nichols, Peter J.
 ; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
 ; NUMBER OF SEQUENCES: 31
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
 ; STREET: Steuart Street Tower, 20th Floor, One Market
 ; CITY: San Francisco
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94105
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US94/01624
 ; FILING DATE: June 25, 1993
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Weber, Kenneth A.
 ; REGISTRATION NUMBER: 31,677
 ; REFERENCE/DOCKET NUMBER: 15280-115
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 543-9600
 ; TELEFAX: (415) 543-5043
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 456 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US94-01624-6

Query Match 32.8%; Score 1306; DB 5; Length 456;
 Best Local Similarity 100.0%; Pred. No. 6.2e-77;
 Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGHGVDGMHVKKEKKNDEKRRKDEERNKTOEHLKEIMKHIVKEVGEAVKKEAAE 93
D 1 AGHGVDGMHVKKEKKNDEKRRKDEERNKTOEHLKEIMKHIVKEVGEAVKKEAAE 60
QY 94 KLLKVPSPDVLKEMTKAIGKTIYVGDITKHSLEALSEDKKTKIDYKGDALLHEHYV 153
D 61 KLLKVPSPDVLKEMTKAIGKTIYVGDITKHSLEALSEDKKTKIDYKGDALLHEHYV 120
QY 154 AKEGEPLVIOSSSDYVENTEKALNVEYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 213
D 121 AKEGEPLVIOSSSDYVENTEKALNVEYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 180
QY 214 SDSGODLLFTNOLKEHPTDSVEFLQNSNEVOEVPKAFAYITIEPOHVDLQYAPEA 273
D 181 SDSGODLLFTNOLKEHPTDSVEFLQNSNEVOEVPKAFAYITIEPOHVDLQYAPEA 240
QY 274 FNYMDKFNEDINL 287
D 241 FNYMDKFNEDINL 254

RESULT 13
5183745-3
PATENT NO. 5183745
APPLICANT: LANCHIN, ANTOINE; GLASER, PHILIPPE; KRIN, EVELYN;
BARZU, OCTAVIEN; LADANT, DANIEL; ULLMAN, AGNES
TITLE OF INVENTION: ADENYL CYCLASE DERIVATIVES AND THEIR
BIOLOGICAL USES
NUMBER OF SEQUENCES: 13
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/426,541
FILING DATE: 25-OCT-1989
SEQ ID NO: 3
LENGTH: 800
5183745-3

Query Match 12.6%; Score 501; DB 6; Length 800;
Best Local Similarity 24.6%; Pred. No. 1.1e-24;
Matches 202; Conservative 145; Mismatches 281; Indels 194; Gaps 36;
QY 4 KKEFT---KVISMSCLVATITLSPVPIPLVQAGHGVDGMHVKKEKKNDEKRRKDE 59
D 3 RNRKIPKRFSTIISVLLFAISSSOAIEVNA-----MNEHTYSDIKRKHKT 49
QY 60 ERNKTQOEHLKEIMKHIVKEVGEAVKKEAAEKLKVPSPDVLKEMTKAIGKTIYVGD 119
D 50 EKNKTEKRFKDSINNVLKTEFTNETLTKIQOTODLKKIPKVDLEIYSEIGETIYFDI 109
QY 120 DITHHISLEALSEDKKTKIDYKGDALLHEHYVAKEGEPLVIOSSSDYVENTEKALN 179
D 110 DLVHKKELQDSEEEKSMNSRGEKVPSPAFVFEKRETKLI-NTKDAINSEOSKE 168
QY 180 VYEIGKILSRDILSKINOPYOKFLD-VLNTIKNA--SDSGODLLFTNOLKE---HPT 232
D 169 VYEIGKISLDITISK-----DKSLDPEFLNLKLSLSDSSDLISQKREKIELNKK 223
QY 233 DVSVEFLQNSNEVOEVPKAFAYITIEPOHVDLQYAPEAFTNYMDKFNEDINLSEEL 292
D 224 SIDINFIKENLTERQHAFLAFSYFAPDRHTVLELAPDMFEYVKNL----- 271
QY 293 KDQKMLSKYER--WEKIKOHQHSNDSLSSEGRGLLKKLOIPIEPKDDIHSLSQEKKE 350
D 272 -----EKGGEFEKI-----SESLKKEG-----VEKDRIDVL---KGEKA 301
QY 351 LKRIQIDSOFLSTEKEFLKLIQIDIRDSLSEEEKELNRIQVDSNPLSEKKEFLK 410
D 302 L-----KASGLVPEHADAFKTI-----ARLNTYILFRPVNKLATN---LI 339
QY 411 KLLKIDIOPYDINORLODTGGLDPSINLDVRYKQY-----KRDIONIDL--HQ--SIG 461
D 340 KSGVATKGLANVHGKSSDMGVPAGYIIPQDLSKKHGOGLAVKGNLKKKSTIEHESIG 399

QY 462 STLXNKIYL-YENMNINNLATTLGADLVDSPDNKINRGIFNEKKNK-----YSISSN 515
D 400 -----KIPKLDHRIEELK-----ENGILLKCKKEIDNCKKYLLLESN 438
QY 516 YMIADINERPALDNERLKRQIOLSPDRAGYLE--NGKILIORNIGLEKDYQ--IIKSEK 573
D 439 NOVVEF--RISDENNEVOYK-----TKEGKITVAGEFNNRNVAKNVEGYKLPLTA 490
QY 574 EYIRDAKVVKSKIDYFIOEAOQLINQENMKALGLPKYTK-----LITFVHNRYA 625
D 491 DY---DLFALPRLSLEIKQIIPQ---KEMDKVNTPLSLEKQGVNMLIKYIERK-- 541
QY 626 SNIVSAVLLINEKKNINQSDLIKVTNYLVQNGREYFDTILPNAIOYTHOD--EY 683
D 542 ---PDSTGKTLNNQKQMLDRL-----NEAVKYTGTCGDVNVNHCTEODNEEPF 587
QY 684 EOVSXGLVPESSSILHSG---PSKGVELRNDSEGFHEGHAVDYA-----GYL---- 732
D 588 EKDNIEFLINPEGEFILTNNEMTGRFLEKNITGKDYLYFNRSYNNIAPGNKAVIEWTD 647
QY 733 -LDRKNSDLVTNSKKFIDIFKEGSNLTSGRTNEAEFFAA 773
D 648 PITKAKINTIFTSAEFI-----KNLSIRSSNNGVYKDS 682

RESULT 14
US-08-973-462-8
Sequence 8, Application US/08973462B
PATENT NO. 6191270
GENERAL INFORMATION:
APPLICANT: DROULHE, PIERRE
TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
FILE REFERENCE: 0660-0125-0 PCT
CURRENT APPLICATION NUMBER: US/08/973,462B
CURRENT FILING DATE: 1998-02-06
EARLIER APPLICATION NUMBER: PCT/FR96/00894
EARLIER FILING DATE: 1996-06-12
EARLIER APPLICATION NUMBER: FR 95/07007
EARLIER FILING DATE: 1995-06-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 1786
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

Query Match 5.7%; Score 228.5; DB 4; Length 1786;
Best Local Similarity 18.9%; Pred. No. 1.2e-06;
Matches 155; Conservative 134; Mismatches 271; Indels 259; Gaps 32;
QY 44 VKEKEKKNDEKRRKDEERNKTOEHLKEIMKHIVKEVGEAVKKEAAEKLKVP--- 100
D 891 IEEKLEELHEVNLVALENTQSEEEKKEVIDVI-----EEVKEEYATLITVEQAE 942
QY 101 -----SDVLEMYKAIGKIY--IYDG--DIRKHSLEALSE--- 132
D 943 EKSANTITEIPENLEENAVESNENVAENLEKLNETVETVLDKYVEYVEISGESLENEM 1002
QY 133 DKRKIKIDYGDALLHEHYVAK--EGEPLVIOSS-----DYVENTEKAL 178
D 1003 DKAFSEIFDVNVKGOENLLGMRSTESTIVIOSEKVDLENVSSILDNINMKKGL 1062
QY 179 -----NVYEIGKILSRD-----ILSKINOPYOKFLDVLNTIK 211
D 1063 LNKLENISSTEGVOETVTEHVEQNVYVDVPAKMDQFLGILNEAGGLKEMFNLDEYFK 1122
QY 212 NASDSGODLLFTNOLKEHPTDSVEFLQNSNEVOEVPKAFAYITIEPOHVDLQYAP 271

Db 1123 SES-----DVTVEEIKDEP-----VQKEVEKETSIEEENIVDL EE 1163
QY 272 EARNYMDK-----NEQETNLSLEELKQDRLSREKMEKIK 308
Db 1164 EKEDLTKMIDAVEESIEISDSKESIKDEKDVSLVEEVDMDSEVEKVELK 1223
QY 309 OHYOH-----SDLSSEGRG-----LKKLOPIE-----PKD 338
Db 1224 NMEELMKDAVEINDITSKLIETOELEVEADLIKMEKLEKALSDSEKIIDAKD 1283
QY 339 DIHLSOEKEKEL-----LKRIOIDS-----SDFLSTEE-----KEF----- 370
Db 1284 DLEKYLEEHDITTLDEVEVELKVDKEDKLEKVDLKDLEEDLIKVEKIKLESEFILE 1343
QY 371 -----LKKLOIDRDSLS-----EEREKLLNRIOVSSNPLSEKEKELKLDI 416
Db 1344 DYKELKTIEDLIEERKEIEKDHFEKEFEAEIKLEADILKEVSSLEVEEKEKLE-- 1400
QY 417 OPYDINORLOD-----TGGLIDSPSINLDVRKQYKRDIONIDA 454
Db 1401 EHELEKEVEHIIISGAHKGLEEDLEEVVDLKGSTLMDLKGDMELGDMKESLEVDVT 1460
QY 455 LHMOSIGSTLYNKTYLKENNNINLTATGADLVSDTNTKINRG-----IFNEFKNF 508
Db 1461 KLGERVESLK-----DVLSSALGMDDEQMKTRKKAQRPKEVLLKEVEKEEP 1508
QY 509 KYSISSNYMIVDINERPAID-----NERLKMRIO--LSPTRRAGYLENGKLLIORNI 558
Db 1509 KKKITTKKVFEDIKDEPKDEIVEEMKDEIDEDIEEDIEEDKEDIEDIEDIEDI 1568
QY 559 GLEIKD--VOIIRKSEKEYIRIDAKVYPKSIDTKIOEAOQUNIOENKALGLPKYKLI 616
Db 1569 G-EDKREVIDLIYQEKERIEKVKAK--KKLEKVEEGVSGLKHVDEVM--KIVQKI 1621
QY 617 TENVHNRYASNIYESAYLLINEMKNNIQSDLIKRYNYL 655
Db 1622 DKEV-DKEVSKALESKNDVTNVLKON--QDFFSKVKNFV 1657
RESULT 15
US-08-480-604A-6
Sequence 6, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KIRK, JOHN A.
APPLICANT: THALEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
FAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-6
Query Match 5.1%; Score 202; DB 1; Length 2710;
Best Local Similarity 20.4%; Pred. No. 0.0001;
Matches 189; Conservative 133; Mismatches 264; Indels 340; Gaps 49;
QY 48 EKNKDENRKRDEERKNTQOEHLKEI-MKHIVKIEVKG--EAVKKAERKLEK---V 99
Db 212 EYNDDEFVLESYRTNS-----LKRINSNHGIDIRANSFLTEQELLINYSOELLNRGMIAA 266
QY 100 PSDYLEM--YKALGKRYI-----VDGDTIKHS-----LEALSEKKK 136
Db 267 ASDIVRLALKNFGG-VYLLVMDLPGIHSDFKTSRPSISGLDRWMKLEALIMKYKY 325
QY 137 IKDIYGRDALHEHYVAKGEYEPV-----LVTSSEDEVENTEKALN----- 179
Db 326 INN-----YTSNFPKLDQDLKDNFKLIIIESKSEKEIETSKLENLVSDLEI 372
QY 180 -VYIEIGKILSRDILSK-----INOPYQFLVLANITKNASDSGDGLFTNOLK- 228
Db 373 KIMFALGSVINQALISKQSGYLTNLVIEQYKNRYQFLNQHLPALIESDNN---FTDTTKI 429
QY 229 EHPIDSEVEFLQNSNEVQEVAKAFAYIE---PQHRVLYQIYAEAF--NYMDKFE 282
Db 430 FHDLSFNSATWENS-----MFLTKIAPYLOVGMPEARSTISLSGGAVASAYDFINL 483
QY 283 QETNL-----SLEELKQDRLSRYE-----KMEKIKOHYOHMSDS 317
Db 484 QENTIEKTLASDLIEKFPENNINSQLTEQFINSUMFSDASAKYQEFKRYVDT--GGS 541
QY 318 LSEEG-----RGLKLIQIP-----IEPKKDI-----IHS 343
Db 542 LSEPDGVDFPNKNTALDKNYLLNNKIPSNVNEAGSKYVHYIIOQDDISYEATCNLFS 601
QY 344 LSOEKEKLLKRIQIDS--SDFLSTEEKEFLKLIQIDRDSLSEKELN----- 391
Db 602 KNPKNSTIIQIRNMNSESXSYFLSDGESITELNKRYRIPERLKNEXKAYVFIFGHGKDEFN 661
QY 392 -----RIQVSSNPLSEKEKELKLDIOP-----VDINORLODPTGGLI 432
Db 662 TSEFARLSVDS--LSNEISFLDTIKLDISPKNVEVNLGCMNFSTDFNVEEYEPKLL 718
QY 433 DS-----PSINDLV--RKQYKRDION--IDALH-----OSIGSTLYNKI 468
Db 719 LSIWDKITSTLPDVNKNKNSITIGANQYEVRIINSEGRKELLASHGKWKKEAIAIMSDLSKE 778

```

QY 469 YLYENMNINLTATLGADLVSTDTNKTINRGIFNEFKKNFRYSISSNYM--IVDINERPA 526
Db 779 YIF-----FDSIDN-----KLKAKSKNIPGLASISE--- 804
QY 527 LDNERLKWRIOQSPDTRAGYIENGKLLIQRNIGLEIKDVQIIOSEKEYIRID---AKV 582
Db 805 -DIRTLLDASVSPDTKF-IINNLKLNIESSIG-----DIYYEKLEPYKN 848
QY 583 VPKSKIDTKIOEAOI--NINOEMNKALGLPKYTKLITFNVHNRVASNIVESAYLILNEMK 640
Db 849 IINHSDLDLIDEFNILEVNSDE-----LYELKIL-----NNLDEKYLISFE-- 889
QY 641 NNIOSDLIKKVTNYLV---DNGRFVFTDITLPNIAEQYTHODEIYEQVHSGLYPES 696
Db 890 ----DISKNNSTYSVRFINKSNNGESYVE--TEKEIFSKYS--EHITKEIST----- 933
QY 697 RSIILHGFSKGVLELRNDESEGFIERGHAVDDYAGYLLDKNOSDLVT-----NSKKFIDIF 751
Db 934 -----IKNS-----ITDVNGNLIDNIQILDHTSQVNTLNAAFIOSL 970
QY 752 KEEGSN---LTSYGRTEAEFFAEAF 774
Db 971 IDYSSNKDVNLNDLSTSVKVLXAOLF 996

```

Search completed: December 2, 2001, 13:49:24
 Job time: 204 sec